The Experiential World

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor of Philosophy by Richard Davnall

24th September 2013

Contents

Introduction

Idealism is the theory that the physical world is ontologically dependent on some facts about the (usually human) mental realm. This is ordinarily taken to be quite a strange position, since we ordinarily take the human mental realm to be constituted by physical things - atoms, quarks or whatever our best physical theories take as their fundamental ontological ingredients. It therefore seems natural to think that all facts about the human mental realm are sustained by physical facts, rather than the other way round.

But the theories of physical science which fundamentally constitute the normal assertion that there is a mind-independent physical world were developed empirically. That is to say, they were formulated based on experience - more precisely, on patterns among our experiences. We ordinarily assume that ontological dependence runs in the opposite direction to epistemic dependence; that our experiences depend ontologically on the postulates of our theories, despite the fact that our knowledge of the theories and their postulates depends epistemically on our experiences.

Idealism, then, could be seen as the attempt to align ontological dependence with epistemic dependence. Were it to be generally seen this way, I feel it would be seen as a much less odd view of the world than it is actually seen as being. The challenge most often levelled at idealism is that of explaining the regularities in our experiences without recourse to physical objects.

And I feel quite strongly that this is a spurious challenge. It amounts to the question of 'why should there be regularities among our experiences?', a question normally answered by the physicalist in terms of physical objects - but the physicalist giving this answer generally conveniently ignores the question of 'why should there be regularities in the behaviour of physical objects?'

Our best physical theories, ultimately, are those which accurately map regularities among our experiences. So in a way, we have already arrived at a form of idealism; facts about the human mental realm (specifically, facts about the character of our experiences) are what make physical theories true or false, or at least better or worse. I intend, though, to go rather further than this.

Specifically, I intend to argue that this putatively epistemic fact - that a 'better' physical theory is one which better captures (describes and predicts) the character of our experiences - is underlain by a metaphysical fact, that our epistemic access to the physical world has an ontologically constitutive role to play in it. Exactly what this rather complex

assertion means must be put aside for now, since explaining it in detail will be the task of the longest chapter of this text, and not easily summarised.

Before we can even begin to tackle this task, a number of preliminaries must be observed. We can identify, broadly, four positions we might take regarding the ontological status of the physical world:

- Dualism[1]: The view that there are irreducibly mental and irreducibly physical entities.
- Physicalism: The view that there are irreducibly physical entities, to which all mental entities in some way reduce.
- Idealism: The view that there are irreducibly mental entities, to which all physical entities in some way reduce.
- Compatibilism[2]: The view that there is a single kind of substance, neither irreducibly mental nor irreducibly physical, in which mental and physical attributes inhere (for some intuitive notion of 'inherence').

The debate which has surrounded dualism at least since Descartes' formulation of it is not one I shall go into here - it is too vast, and since dualism preserves the ontological fundamentality of the mental realm anyway, I regard it as rather less of a threat to idealism than the other alternatives. The main controversy I shall discuss is that between physicalism and idealism, but first I shall spend some time on the view I have characterised as 'compatibilism'.

Compatibilism, I feel, turns on an awkward and counterintuitive use of the terms 'mental' and 'physical'. Dualists, physicalists and idealists are often unclear about exactly what these terms mean, either offering only vague, ostensive definitions or leaving the matter to the reader's intuition. In chapter 1, I attempt to get clearer about what 'mental' means in ordinary language, concluding that anything which is subjective, but not (truly) quantifiable or (truly) spatial, is mental.

In chapter 2 I examine the links between these prima facie dissociated properties. I demonstrate that anything which is subjective is necessarily non-spatial and non-quantifiable, and furthermore that anything which is non-spatial or non-quantifiable must

---

[1] Technically, this is substance dualism only.

[2] I am using 'compatibilism' here as a blanket term covering a wide range of very diverse positions, including but not limited to Russellian monism, panpsychism and property dualism. While this breadth would in most instances be problematic, I indulge in it here only because I believe the same argument - the one I shall present in the first two chapters of this text - applies to all these theories.

turn out to be subjective. I go on to show that this entails that 'mental' and 'physical' are both mutually exclusive and mutually exhaustive terms - that nothing can be both, but everything is either one or the other.

I take much of this to be implicit in our ordinary way of thinking about minds – that we cannot read each other's minds (they are not public), that scientific measurement is only possible of brains, not minds, and that we cannot point to a thought. If correct, this entails that any theory which denies that 'mental' and 'physical' are mutually exclusive and exhaustive is revisionary. Either way, it seems to me that a definition of the mental as essentially subjective, qualitative and non-spatial, and thus mutually exclusive with the physical, makes a good standard or conventional definition. Compatibilism, by definition, opposes this, by insisting that there are (at least) some objects which are either both mental and physical, or neither. Any compatibilist position, I argue, will turn out to be a rebranding of either physicalism, idealism or dualism.

In chapter 3, I move on to consider what has, historically, been one of the most popular strategies used to argue in favour of idealism. This is what I have called the *spatial* strategy. If the definition I have used is correct, then the physical world must be spatial; the spatial strategy argues either that spatiality is an illusion or that space is constituted by facts about our experience. The versions of this strategy that I consider in detail are those of Kant and of the British idealists, primarily Bradley. I conclude, however, that their argument has a significant flaw, in that they assume that our experiences furnish us with more generous epistemic resources in respect of what is ontologically fundamental than they actually do.

From this, in chapters 4 and 5, I move on to discuss a more sophisticated version of the Kantian strategy developed by John Foster. Here, the key point is that most physicalists and idealists would agree, I think, that what I will call the *experiential* world - the physical world as we experience it, consisting of common-sense macroscopic objects like tables and chairs - is *constituted* by some underlying, fundamental world. Foster's argument is that the relationship of constitution between the fundamental world and the experiential involves some mental facts - and that this is a less trivial statement than it seems.

The precise details of this will have to wait until chapter 5 (chapter 4 focuses on a key, complex preliminary step). I said that most physicalists and idealists would agree that common-sense macroscopic objects are constituted by some underlying, fundamental things (whatever they may be). There are positions on both sides that would not, however. On the idealist side, there is pure phenomenalism, which says that the phenomenal

qualities of these objects are themselves ontologically fundamental (and that nothing else is). The most common argument for this view is one of absolute parsimony, and while I have some sympathy for this argument, it seems to convince very few, and I do not intend to discuss the matter further in this text.

On the physicalist side, the equivalent position is naive (or direct) realism, which takes the apparent properties of macroscopic objects to be both physical and fundamental. In chapter 6, I offer what I feel to be the strongest of many different arguments against this position. Taking naive realism to entail that whatever can be veridically perceived or logically inferred from perception must be true[3], I show that, thanks to the special theory of relativity, naive realism must entail a contradiction.

In combination, I take these arguments to show that the experiential world has at least a partial dependence on some mental facts. This is a rather weak form of idealism, but arguably it lays the groundwork for a defence of a stronger position. Such a defence is beyond the scope of this text, however.

[3] I recognise that this is a very strong form of naive realism, but I will argue that weaker forms fail to escape Foster's argument as presented in chapter 5.

1. Hallmarks of the Mental

How are we to understand the terms 'mental' and 'physical'? How would a *philosopher's* dictionary define them? Few physicalists, idealists or dualists have ever agreed on more than a very basic, paradigm-based definition ('the pen is physical; my experience of and thought about it are mental'). This clearly falls far short of what we need.

In a landmark essay originally published in 1958, the compatibilist Herbert Feigl identified eight putative hallmarks of the mental. His intent was to go through these one by one and show that for each hallmark, there were either some archetypally mental phenomena[4] which didn't possess it or some archetypally physical phenomena which did. The essence of Feigl's argument is that none of these hallmarks identify an exclusive mental field in line with our intuitions, and thus 'the mental' should not be regarded as a separate realm to 'the physical'.

I think Feigl is correct in this strategy as far as it goes; that is to say, 'the mental' is not wholly captured in terms of any one particular attribute or feature. However, Feigl gives no thought to the possibility that two or more of his signs might cumulatively define the mental realm. I will return to this point later; first, I should introduce Feigl's classification. In doing so, I will identify some problems with the way he draws the various dichotomies involved.

Feigl presents each hallmark as a dichotomy between one attribute ordinarily assumed to be mental and one assumed to be physical. His dichotomies are:

- Subjective/objective
- Non-spatial/spatial
- Qualitative/quantitative
- Purposive/mechanical
- Mnemic/non-mnemic
- Emergent/compositional
- Holistic/atomistic
- Intentional/non-intentional

Some of these distinctions are obvious and familiar; some also have obvious flaws ready for Feigl to exploit. Certainly, I do not think all eight will be of use to us. For starters,

---

[4] I use 'phenomena' here and throughout what is to come in a non-technical sense, meaning simply 'things that occur/exist/are part of apparent or fundamental reality'.

the debate over intentionality is far too vast, too clouded with competing definitions and models, for us to hope to make real headway with it in the space available - to do so would take a whole book by itself. Fortunately I think we can develop a satisfactory and intuitive definition of 'the mental' without introducing the notion of intentionality at all.

Feigl is dismissive, too, of the emergent/compositional and holistic/atomistic distinctions. He treats the two dichotomies as equivalent, and I confess I can see little difference between a thing's being emergent and its being holistic, or being a composite and being atomistic. Feigl's complaint is that there are plenty of non-mental things (water being the perennial favourite example) which are emergent - one can even argue that all macroscopic physical objects emerge from the chaotic jumble of subatomic particles which physical science now posits as their building blocks[5].

Mnemicity and Purposivity

'Mnemicity' is perhaps the most unfamiliar of Feigl's terms. It means, in essence, the ability to learn from experience. This is certainly something we tend to associate with mental activity - for example, we hold animals that can learn to be more intelligent than those that cannot, and we hold a lack of ability to learn in a child, generally, as evidence of a mental disorder (whether or not this diagnostic approach is justified) - but it is perhaps a less intrinsically mental phenomenon than some of those remaining to be discussed.

Something similar goes for purposivity, the ability to act with a goal in mind and make plans to reach that goal. When, for example, we see a crow dropping stones into a glass of water to raise the water level to the point where it can reach to drink, it is the clear intent and planning which makes us attribute intelligence to the crow, but do we, at the same time, attribute a full-blooded mental life to the crow?

My point is that mnemicity and purposivity are attributes we take as *empirical* hallmarks of the mental without necessarily normally taking them to be definitive of the mental. We use evidence of purposive activity and the capacity to learn to judge whether other creatures have minds, but it would be, I think, quite strange to answer the question 'Am I, myself, minded?' in terms of the ability to demonstrate learning or purpose.

Crucially, our definition of 'mental' must give us a useful sense of the term 'mental state', and individual mental states clearly are not always either mnemic or purposive; one

[5] While it's true that this emergence is normally much more transparent than it is in the case of the mental, for the time being I am happy to let Feigl's objection stand; again, we do not need these distinctions for our definition of the mental.

could hardly attribute the capacity for learning to a purely perceptual state, and it is hard to see how a single stray thought, in isolation, could be considered purposive. We may quite legitimately require that *minds* be mnemic and purposive - this, I think, is the implicit intent of the question being asked about the 'intelligence' of a crow - but it seems very odd to require every mental state to meet these criteria.

Mnemicity and purposivity are key components in the debate about 'other minds', to be sure, but our question is not 'Under what circumstances can we attribute a mental life to a creature?'; it is 'What does it mean to say a thing is 'mental'?'. And I hope I have shown that this is a question that mnemicity and purposivity cannot answer. We can now turn to the remaining three of Feigl's hallmarks, all of which I feel are rather closer to what we need.

Subjectivity

Opinion

Feigl identifies three different senses in which he thinks someone might take 'subjectivity' to be the hallmark of the mental. Of the three, the first and third I think are two different misconstruals of the same point (which I will come to in a moment) and the second is merely obviously wrong. The second sense Feigl invokes is the sense in which 'subjective' is opposed to matters of 'objective fact' i.e. subjectivity as meaning 'opinion'.

This is an entirely legitimate understanding of the term 'subjective', but I do not think any mental realist has ever held it to be the hallmark of the mental. If nothing else, mental realists generally look quite kindly on the existence of non-propositional mental states (raw experiences, for example) which clearly are not capable of being either matters of fact or of opinion. This is to say nothing of dispositional and subconscious states.

Even if we broaden the dichotomy a little, to differentiate between the assertions 'it's cold outside' and 'the temperature outside is ten degrees Celsius', we would be no better off. While the former assertion is certainly about something subjective - the way I feel, or rather felt at a particular time, in an environment of a particular temperature range - and the latter an assertion of an objective fact about that environment, it is unclear how this distinction could be generalised to cover mentality in any systematic way.

In a sense, then, Feigl is right to reject this meaning of 'subjectivity' as a hallmark of the mental, but he is wrong if he thinks that he has scored a victory against mental realists in doing so. Something similar goes for the rest of his treatment of subjectivity-as-hallmark-

of-the-mental, which focuses on the idea of privacy ('privileged access' might be a better term).

Privacy

I have already said that I believe Feigl is talking about the same thing in his first and third interpretations of 'subjectivity'. Before getting to that problem, though, I'll summarise Feigl's treatment as best I can. Again, Feigl is attempting to show that privacy/privileged access cannot be the essence of mentality because it is not mutually exclusive with its opposite (objectivity, for some publicity-based definition of 'objectivity'), which he presumes to be essentially associated with physicality. That is to say, he is trying to show that there is at least some sense in which a 'subjective' state is also objective.

Here is the essence of Feigl's treatment:

> *"To say that a twinge of pain experienced by person A is 'subjective' or 'private' to him may simply mean that another person B, observing A's behaviour, may* infer *A's pain, but does not* have *it, i.e. he does not directly experience it." (1967, p.30)*

We can, I think, get quite a good intuitive hold on the way Feigl has drawn the 'infer/have' distinction; 'infer' simply means 'know about'. And this is exactly Feigl's argument; specifically, he takes seriously the claim that B can know about A's subjective experience. I am quite happy to accept that for one to know (something) about an object, it must be in some sense objective, and I think this is a fairly reasonable condition to apply to any definition of 'knowledge'.

Feigl's treatment of this sense of 'objective' takes it to mean 'in-principle-intersubjectively-confirmable'; that is to say, he takes it that for something to be objective(ly true) is for there to be some possible (even if utterly implausible to execute) test by which its truth can be established. This verificationist conception seems more than a little out of place now, but Feigl did admit that he was still emerging from the influence of positivism (as was philosophy in general, in the late 50s when he originally wrote the essay).

By way of illustration, Feigl offers the example that it seems natural to say that 'President Eisenhower experienced intense pain during his heart attack' is an objective fact - that is, it is in principle intersubjectively confirmable - even though the pain is itself subjective. Thus, argues Feigl, Eisenhower's pain is both subjective and objective, and so, if subjectivity is mentality and objectivity is physicality, compatibilism must be true.

Before we get onto the general problem with Feigl's approach to subjectivity, there is a serious problem with the approach just described. Examining it will bring out the confusion Feigl makes between this looser sense of subjectivity and his final sense of it, which is the stricter 'absolute subjectivity'.

Variable Realisability

The first part of this problem should be obvious; Feigl claims that Eisenhower's pain during his heart attack is intersubjectively confirmable (at least in principle), but he gives no account of how this is to be achieved. This would not be a problem if his assertion of the intersubjective confirmability of subjective states were uncontroversial (and indeed, Feigl seems to think it *is* uncontroversial), but to assert this is in fact to paper over one of the deeper controversies between the mental realist and the physicalist.

Feigl's claim of how we are to understand this confirmability is steeped in post-war optimism about the scientific project. He says:

> *"Modern devices, such as the lie detector, and various clinical-psychological techniques enable us to test for such 'private' events with increasing (though generally only relatively low) reliability." (1967, p.32)*

The implication of the 'increasing (though generally only relatively low) reliability' here is clearly that Feigl expected the reliability of these devices and techniques to keep increasing. I don't know what he had in mind by 'clinical-psychological techniques', but the 'lie detector' has come under increasing fire in the last decade for generating an unconscionable rate of 'false positives' - accusing people of lying when they are not. (National Research Council, 2003, p.214)

While Feigl's optimism may have seemed justified in 1958, the intervening half-century has taught us time and time again that the mind is just too complicated to be measured or tracked in anything like the way envisaged. Let us give Feigl the benefit of the doubt, then, and assume that he would have reformulated his position in light of scientific progress. What might he now claim to be the method by which mental states are intersubjectively confirmable?

Let us put aside any questions of telepathy[6] and take Feigl as claiming that Eisenhower's pain is identifiable by means of some intersubjectively observable phenomenon. Care is needed in how we take 'identifiable by means of' in the previous sentence; obviously, it would be easiest to take it in the strongest philosophical sense, as meaning 'numerically identical with', but this may be too strong. It might be that Eisenhower's pain is held not as *identical* with whatever observable we use to confirm it, but as necessarily one-to-one correlated with it (i.e. the existence of one always implies the other, and vice versa, and nothing else does).

This weaker sense of identification allows that we might identify Eisenhower's pain-state with a neurophysiological state (say, the high-intensity firing of a large number of pain nerves around the heart) without being immediately refuted by the obvious qualitative differences between the two[7]. The problem runs rather deeper than this, however, and it has to do with whether it is possible to learn about correlations between neurophysiological and mental states.

Whether or not there *is* a correlation between neurophysiological and mental states, if it is impossible to learn about such a thing then we cannot ever know it exists. We certainly cannot use it to make inferences from neurophysiological states to mental states - which is what we are taking Feigl to have envisaged. In short, if it is impossible to develop knowledge of the envisaged correlation, then we will have no mechanism for confirming the mental states of others.

I am going to take it for granted that pain, like all other immediate experiences, is sui generis and known only by association. That is to say, one can only understand the proposition, 'Eisenhower felt pain' if one has oneself felt pain. This has always seemed to me the only plausible way to understand qualitative sensory language, though I am aware

---

[6] It is not even clear that we would be entitled to treat the relevant, strict form of telepathy - the sharing of a single mental state, rather than the acquisition of a mental state qualitatively similar to that belonging to another subject - as a logical possibility. At very least, for Feigl to do so (and he shows no such inclination) would be to beg the question, since it would imply that a mental state could be full-bloodedly non-private.

[7] I think the same cannot be said of the attempt to identify Eisenhower's pain with any of his pain-indicating behaviours. After all, for his heart-attack-related pain to be intersubjectively confirmable, we must be able to distinguish it from other intense, systematic pains Eisenhower might be having (a serious spinal condition, perhaps, or a severe case of poisoning). A strict behaviourist might be able to argue that all these conditions actually produce the same pain state and thus the one-to-one correlation between behaviours and pains is preserved, but I very much doubt this strategy could be pursued far enough to do justice to the full richness of the mental realm. By way of example, consider whether any behaviour might distinguish between my deciding not to steal money from a wallet left unattended because I believe it to be wrong and because I fear reprisals if caught.

the position is not without controversy. Still, the only common element between all the pains I can remember feeling is pain, and much the same goes for colours, scents, sounds etc.

So, if I am to confirm Eisenhower's pain neurophysiologically, I not only need to know about the correlation between his pain state and some firing of his nerves, but also about the correlation between his pain state and some pain states of mine, which will presumably be accomplished by way of my neurophysiological states. So, I need to know how my pain-state relates to my nerves firing, and how my nerves' firing relates to Eisenhower's.

And therein lies the problem. At the level of individual cells, the layout of my nervous system may be very different to Eisenhower's. He might have half as many nerves in his heart area as me; I might have had part of my brain removed as a result of a serious illness; I might have had my corpus callosum severed as a treatment for epilepsy. The point, put less drastically, is that I may not even be capable of the same pattern of nerve firings as sustains Eisenhower's pain-state, and even if I am, the resulting state might feel very different to me than Eisenhower's does to him.

To say that in such a case Eisenhower and I do not have relevantly similar experiences is to miss the point. The point can be made more generally, though, by going a little further afield. Let us assume that at some point in the future, mankind splits into two separate species, humans and humanas. One of the ways humanas differ from us is that their nervous systems operate on a completely different biochemistry which, instead of involving a complex network of cells, involves a single large cell spread throughout the body. Humanas exhibit every sign we could wish for of human-style mental states, to the point that it becomes absurd to not attribute mentality to them, but there is absolutely no possibility of a humana's physiological state structurally resembling a human's.

The general point, then, is that mental language involves predicates which are *variably realisable* - they can, at least in principle, apply to several different kinds of physical system. That is to say, it is not generally possible to correlate the states they pick out on a one-to-one basis with physical states. And given this, we must accept that it is never possible to give an exhaustive specification of all possible physical states which might correlate to a given mental state, since we can never be sure that we have found all the ways in which physical states might realise the mental state in question.

Nor can the compatibilist escape by saying that I don't need to have experienced a pain like Eisenhower's, only a generic pain, to enable me to use the assertion 'Eisenhower

felt pain'. If this were acceptable then we could start to learn about a correlation by identifying humana-pains behaviourally and tracing them to humana neurophysiological states. Unfortunately, this approach cannot go far enough.

For Feigl's objection to succeed, there must be no part of the mental which is not in-principle-intersubjectively-confirmable[8]. So it's not enough that we can confirm *that* Eisenhower felt pain; we need to be able to confirm everything that is distinctive about Eisenhower's pain, and thus we need to know the correlation between Eisenhower's pain-state and his neurophysiological state in detail. Any mismatch between the level of detail allowed by the correlation and the detail of Eisenhower's pain-state constitutes an acknowledgement that there is some entirely private aspect to that state.

This brings us neatly to the final way in which Feigl understands subjectivity, and also reveals the basics of his confusion between the two. The point here is that there are some elements of mental states which no method of intersubjective confirmation could tell us anything about. Feigl refers to these elements as the 'absolutely subjective' part of the mental.

Absolute Subjectivity

Feigl takes the absolute definition of subjectivity to be a separate sense from the privacy-based version, though perhaps not a completely separate sense. His definition of the absolutely subjective is simple: anything which is *not* in-principle-intersubjectively-confirmable is absolutely subjective. His argument will be that it is (scientifically) meaningless to assert the existence of anything which is absolutely subjective.

Feigl's objection to absolute subjectivity, then, is that the assertion that some phenomena are absolutely subjective can never be meaningful. Here again, his lingering verificationist leanings come through. He takes intersubjective confirmability as a condition of scientific meaningfulness, which is fair enough, and then takes scientific meaningfulness to be the only criterion of meaningfulness in which we should be interested. This latter step will prove problematic later, but his arguments on the former bear some exposition.

---

[8] It might be argued here that where physical systems are significantly similar - as mine and Eisenhower's may be taken to be, while mine and that of a humana might not - we can generalise from introspection, from our own cases. I'm not convinced that this is a helpful suggestion, since it seems to limit the ability of our language to refer to the mental states of non-human organisms, but even if that's not a problem, I don't think a compatibilist pursuing this strategy could give a clear account of what counts as sufficiently similar.

In essence, Feigl argues that the assertion of absolutely subjective states requires a private language to be meaningful. A subject can only assert the existence of his absolutely subjective states in a language which can refer to them; but only the subject can refer to his own absolutely subjective states, so that language must be private. Feigl's objection is not simply a 'ruling-out' of all private languages; instead he claims that there is no sense to the idea of a private language such as the one just envisaged.

I am a little unclear on the fine points of Feigl's strategy, and he does not draw it to a full conclusion himself, but the clearest point I could find in his discussion is the claim that nothing could render the terms of a private language unambiguous: he asks, "Am I to describe the way I feel at a given moment as 'happy', 'joyous', 'merry' ... or what?" (1967, p.38) We might take this in two ways.

First, we might take Feigl to be asking 'how can I tell which of the universals my private language furnishes me with is the one under which my current state falls?' This is the obvious way to take the question, but if it is what Feigl means as an objection, it is also obviously misguided. The whole point of a private language of the absolutely subjective would be that each universal (or predicate) would have a clearly-defined field of application. 'Happy' and 'joyous' are ambiguous (at least, if we take them to be qualitatively distinct from one another at all) only in virtue of the fact that they belong to an intersubjective language.[9]

The second option is a little stronger (and, in Feigl's defence, some of his remarks in paragraphs preceding the above quote suggest he may have been leaning in this direction). We might take Feigl to be asking 'given that I have been furnished with a number of private language universals by acquaintance with their instances, how am I to tell which my current state falls under? How am I to match it to one or more of my previous states?'

Let us imagine that an hour ago, I had an experience of something rather like happiness, but more energetic; an experience which did not fit any of the classifications already existing in my absolutely subjective language. I introduced the predicate 'joyous' to my absolutely subjective language to mean 'this feeling'. Now, assume I have a repeat of

---

[9] By this I mean that I can make a precise use of the term 'joyous' if I feel there is something qualitatively distinct about it that picks out my present state as similar to past states which I have associated with the predicate 'joyous'. However, if I describe my state to you as 'joyous', your understanding of the term will relate it to past states of *yours* which are likely to be different to mine, if in fact you have any reason for feeling that 'joyous' picks out something qualitatively distinctive at all.

the feeling at the present moment. We can take Feigl as asking 'how am I to know that my present feeling fits under my new predicate 'joyous' as opposed to some alternative?'

I can only answer that the imprecisions of memory that make the question significant are entirely familiar from every other circumstance in which we might wish to apply a predicate. This is such an obvious point that I can only assume I've missed something crucial in Feigl's argument that distinguishes the private language case from the normal version, but for now I shall take the point as settled.

Privileged Access

As we saw at the start of the previous section, Feigl's 'absolute subjectivity' is just a narrowed form of the privacy-based definition of subjectivity discussed earlier. An item is absolutely subjective if there is absolutely no possibility of intersubjective confirmation of it; we might say that an item is absolutely subjective to a given subject if that subject has *privileged access*[10] to it. The key structural problem with Feigl's argument, as I see it, is that privileged access is surely the essence of what the subjectivity-mentalist was getting at all along.

That is to say, if we were to survey philosophers who hold to the contention that the mental is the subjective and ask what they meant by subjective, I would be very surprised if the majority did not outline some notion of subjectivity which boiled down to privileged access. Specifically, I think we would find them holding to the contention that while the existence of mental states is intersubjectively confirmable in some broad sense, the qualitative features of those states are privileged to their individual subjects[11].

This position falls neatly into the gap between the two versions of subjectivity-mentalism that Feigl considers. It circumscribes a particular group of phenomena - the absolutely subjective - as wholly distinct from the physical, but still allows for the possibility of meaningful existence claims about these phenomena. That is, it does so if it can be unpacked in a way which does not fall foul of Feigl's arguments against the positions it supersedes.

---

[10] Where 'access' is defined in some generic but fairly intuitive way that includes but is not limited to perceptual access. One could almost substitute 'interactions' for 'access', but this may open the door a little too wide.

[11] Perhaps this is too strong. Many mental realists might allow that the qualitative features of privileged states are intersubjectively confirmable 'in some broad sense', too. For example, it might be held that if one stands beside a colleague whom one knows to have no physiological or neurological abnormalities and looks in the same direction, one knows that, broadly, they are having qualitatively the same experience as oneself. Assuming they are not a zombie, at least.

Since Feigl's argument against the broader, privacy-based subjectivity-mentalism amounts to the claim that it does not adequately separate the mental and the physical, we can pass it over. Our revised form of subjectivity-mentalism surmounts this by invoking absolute subjectivity. The problem remaining, then, is whether the assertion that there are states to which subjects have privileged access can be rendered meaningful.

Feigl's objection was that any assertion concerning an absolutely subjective state would have to be rendered in a private language. This is true if there is no conceptual separation between the existence of a state and its particular features, but I don't think this condition can be met in any intelligible way. I can quite happily attribute a pain-state to Eisenhower without knowing any of the precise, distinctive features of that pain - whether it is an ache or a burn, whether it spreads throughout his body or is narrowly localised, how severe it is and so on.

The reason this claim does not fall foul of any of the objections I brought against Feigl in my discussion of variable realisability is because the attribution of a mental state is a much more general assertion than those I took Feigl as claiming to be possible. That is, I took Feigl to be claiming that all features of the mental realm could be captured by some system of (putatively physical) observables, which would require us to be able to correlate mental phenomena on a one-to-one basis with physical phenomena.

Our modified version of absolute-subjectivity-mentalism requires no such level of detail. We can retreat to the everyday position of attributing mentality on behavioural and neurological grounds, with the caveat that the qualitative details of the attributed mental states are going to be privileged, i.e. not intersubjectively confirmable. On this reading, 'Eisenhower felt pain' means simply 'Eisenhower exhibited pain-indicating behaviours and certain distinct groups of his nerve cells fired'[12].

In a way, then, our assertion of Eisenhower's pain-state is very different in meaning to what Eisenhower himself would mean if he said 'I am in pain'; it refers primarily to different features of the world. This is unproblematic, though, since it is nothing more than the ordinary position we find ourselves in in respect of the problem of other minds. We

---

[12] This should not be taken to imply that the intersubjectively-confirmable phenomena cannot be reduced to mental phenomena; nor should any of the foregoing be taken to exclude the possibility that absolutely subjective phenomena can be ontologically unpacked in some fundamentally physical way. My point at this stage is simply that the terms 'mental' and 'physical' are mutually exclusive; if there is only one kind of substance, then it must be either mental or physical.

ordinarily evaluate the mentality of others by a very different metric to the one we use for ourselves.

The Subjectivity of Unconscious States

Feigl has one last argument against taking subjectivity to be a (or the) hallmark of the mental, and in some ways it is his strongest. Ultimately, though, I believe it can be surmounted. His argument is that there are some states we would paradigmatically consider to be mental which are quite definitely not subjective. The states he has in mind are subconscious states, though perhaps not all subconscious states[13].

The argument is that consciousness is an essential part of the privacy that Feigl takes to be constitutive of subjectivity. Because subconscious states are hidden even from the subject himself, they are (technically) private to no-one and thus not subjective according to Feigl's definition. I agree with him this far; certainly, we want to take subconscious states to be mental (it would be very strange, though not completely implausible, to suggest conscious states are mental and subconscious ones not), and there is no question that subconscious states stand in a different relationship to their subjects than do conscious ones.

However, the fact that we *do* stand in some relationship to our subconscious states offers an escape that Feigl apparently failed to consider. In our final understanding of subjectivity, we have invoked the notion of 'privileged access' rather than merely 'privacy'. My access to my mental states, whatever form it takes, is privileged if only I can have it. It is the privilege, not the access, that is the key to subjectivity as we have come to understand it.

There is nothing to stop us saying that 'access' is a category term, encompassing several different kinds of relation. This would allow that while we access our subconscious states in a different way to our conscious states, that access remains privileged in the right way to constitute subjectivity. We might even say that our access to our subconscious states consists in *being affected by* them; so long as it is completely impossible for anyone else to have that access, the requisite privilege is maintained.

---

[13] Arguably, there are some neurological states - autonomic reflexes, visual pre-processing and so on - which are part of our subconscious without being mental, but the only reason for this classification would be if we were taking all neurological states to be either conscious or subconscious - that is, disallowing the idea of a 'non-conscious' neurological state. And I cannot see a good reason for this prohibition.

It may even be sufficient to take 'access' in the broadest possible sense, as simply meaning 'standing in some immediate relationship to'. You can only relate to my mental states through me; I can relate to them immediately. My access, my relationship to them, is privileged because of exactly this fact.

I can think of no relationships that are privileged to one or both relata in this way that I would be unhappy to say pick out mental phenomena. There are certainly relations between categories which are in a similar sense privileged to those categories - as a simplistic example, we could take child-bearing to be privileged to female organisms (by making this a feature of our definition of 'female') - but we can simply add to our privileged-access-subjectivity definition of the mental that the access in question must be privileged *to an individual*.

There is, of course, also a modal concern here. It might sound, when put this simply, as if I am committed to the idea that all my actions are privileged to me, since actions are putatively individuated partly by being the actions of a particular agent - my actions are necessarily mine, and someone else standing in the same place and performing a qualitatively similar action would nevertheless be performing a numerically distinct action of their own. My own feeling, though this is a debate too complex to get into here, is that this is too stringent an approach to the individuation of actions, and that someone else could have performed *the same* action as the one I actually performed. This does not apply to mental states, however, since in at least some important respect I *am* the collection of my mental states.

In this way, I believe we can show that any particular item must be either mental - (some kind of) access to it is privileged to some subject or other - or physical - no kind of access to it is privileged. This is a positive start, but much more remains to be said.

Space

We turn next, then, to space and spatiality. I will take it as a basic datum that a thing is either spatial or non-spatial[14], and cannot be both. My contention is that by 'the physical' we must understand something intrinsically spatial, and by 'the mental', something intrinsically not.

Feigl disagrees:

[14] There may be different ways in which a thing or set of things can be non-spatial, though. For example, we might want to differentiate between 'truly non-spatial' abstracta like 'love' and 'justice', and 'pseudo-spatial' abstracta like the set of all triples of numbers, which can of course be construed as a set of spatial coordinates.

> *"I am perhaps not too acute in matters of phenomenological description but it does seem to me that my feelings and emotions pervade large parts of my body-as-I-experience-it ... In the phenomenal field of the subject, specific feelings may be located at least vaguely or diffusely in some not very sharply delimited part of the organism. My feelings or sentiments of elation, depression, delight, disgust, enthusiasm, indignation, admiration, contempt etc. seem to me to be spread roughly through the upper half or two-thirds of my body."(1967, p.40)*

The contention here is presumably that sensations can be associated with the locations at which they are sensed, and thoughts, by analogy, associated with the brain that entertains them[15]. As a first objection, we can point out quite simply that sensations often seem to move when we try to pin down their location. While it's true that our sight stays more or less where our eyes are, I'm sure we've all had the experience of going to a doctor and being unable to say exactly where a given body part hurts (or, failing that, of reaching to scratch an itch and finding the itch to never quite be where one scratches).

To render this objection precise, we must clarify exactly what it is that we mean by spatiality. After all, it is certainly true that my phenomenal field *seems* spatial. My position therefore must be that this spatiality is not the spatiality we're interested in. So what is the spatiality that we're interested in?

When a mental realist says that the mental is non-spatial, I don't think he necessarily has to be saying that mental phenomena never exhibit spatial or pseudo-spatial behaviour. There is a less stringent claim which can be more easily defended; this is the claim that mental phenomena are not occupants of *real* space. That is to say, mental phenomena may exist in some space - the space of the phenomenal field, for example - but they do not exist in the particular space occupied (or, if we are to be Leibnizians, constitutively sustained) by physical objects.

Of course, any full-blooded idealist is eventually going to have to make this claim anyway, since on a final idealist analysis, physical space and physical objects must turn out to supervene on mental objects (and possibly also mental spaces). Indeed, as we shall see in chapter 3, historically the attack on the reality of physical space has been a preferred

---

[15] From Feigl's point of view, of course, all that is needed is that some mental phenomenon or other turn out to be spatial. If we are to construct the robust definition I seek, I must prove this impossible.

strategy of many idealists. I do not intend to pursue that attack here, though; for the time being, let us take as neutral an ontological position as we can and examine the *apparent* world.

The apparent world certainly includes both mental and physical phenomena. The physical phenomena, I think, can be taken as axiomatically spatial. The question is whether the space to which they belong also plays host to some mental phenomena. And, assuming that direct realism is false (which I shall attempt to prove in chapter 6), it seems obvious that it does not.

If direct realism is false, then our experiences are only representationally related to their objects. Information about the objects travels by various physical processes and in several distinct streams through our sense organs and brain and is at some point converted into phenomenal data and distributed throughout our phenomenal fields. The mind cannot be held to be entirely passive in this process, since distortions of the mind such as intoxication, illness and delusion can affect the resulting distribution.

So 'real' space and phenomenal space are numerically distinct - the phenomenal space of a subject is not a part or sub-region of real space. But surely this is a facile point - after all, the correspondence between phenomenal space and real space is good enough that we can navigate real space purely based on the positions of things in phenomenal space. Can't this correspondence be used to establish a schema for locating mental phenomena in real space?

The question of precision is paramount here. An object that is actually in real space, it seems, should have a definite position in that space (though probably at this point we need to stop thinking of macroscopic objects and change our focus to their subatomic building-blocks). Any given particle is present at some spatial point or set of points, and absent at all others.

If we are to make the locations of mental phenomena precise in this way, first we need to establish how to locate them at all. That is, how do we take a technically vague statement like 'It hurts here' or 'my thoughts seem to be located mainly just behind my eyes' and identify which points of real space it applies to?

I can see no plausible way of doing this other than neurologically. But this approach brings up too many unanswerable questions. Nerve cells, though small, each contain countless subatomic particles and are spread over a vast number of spatial points. The neurological method for locating mental phenomena thus has a finite resolution, whereas

space either has infinite resolution or a maximum resolution far greater than the human nervous system (a debate too complex to get into here).

So we can ask whether a thought is present throughout the entirety of a nerve cell, at every point within the boundary of the cell (if, at the subatomic level, the boundaries of a cell are clear in their own right). And if not, which parts of the cell contain parts of the thought? Either way, which parts of the cell contain *which* parts of the thought?

Of course, such questions are partly rhetorical flourish. The point, ultimately, is this: if the contents of the phenomenal field were in some way located in real space, there would have to be a one-to-one correlation between phenomenal field-points and points in real space, but our phenomenal fields simply can't have enough points to match this demand (whether or not locations within our phenomenal fields are in any way precise in themselves).

There is a legitimate question, too, over *extension*. Objects that exist in space, we generally assume, are extended through space; they have dimensions as well as locations. So if there are spatial mental phenomena, we should be able to answer questions about their size.

So, how big is my experience of the view from my window? It's not clear that this is a meaningful question at all. It's certainly not as big as the physical window, since the experience putatively fits inside my head, but the window is far too large to do so. How might we measure the size of an experience? This is the question I shall tackle in the next section, where we shall move on to Feigl's criticism of the qualitative/quantitative distinction. I hope there to make clear in a more systematic way the problem with claiming that mental phenomena are ever spatial.

The Qualitative/Quantitative Distinction[16]

Feigl's Argument

Feigl attributes the following claim to some mental realists: the mental is distinguished from the physical by the fact that mental judgements are qualitative whereas physical judgements are quantitative. That is to say, 'such-and-such a colour is blue' is a judgement about something mental, whereas 'this electromagnetic wave has a wavelength of 460 nanometres' is about something physical.

[16] In what follows, I apologise for the brutish grammar of the predicates 'qualitativity' and 'quantitativity'. 'Qualifiability' seemed little if any better, and its analogue 'quantifiability' has a formal-logical significance I judged it better to avoid confusion with.

To be precise, a judgement is quantitative if it involves a numerical measurement or assertion. Quantitative judgements are about *amounts* of things and are generally subject to mathematical analysis. Qualitative judgements are harder to pin down, except by saying that they are the opposite of quantitative (which in this case would be question-begging). It may be enough for now to say that a qualitative judgement is one about the nature of an attribute rather than its quantity or magnitude.

Feigl's strategy will by now be familiar; he argues that one cannot identify the mental with the qualitative because there are mental phenomena which can be judged quantitatively and physical phenomena which can be judged qualitatively. The latter is the simpler claim to deal with, so I shall start there.

The claim as I understand it is this: while physical magnitudes are measured quantitatively, there are several different *kinds* of physical magnitude, among which the distinctions are qualitative. For example, electromagnetic waves differ qualitatively from masses and again from temperatures.

There is certainly something to this claim, as one cannot measure the temperature of an electromagnetic wave, or (prima facie at least) the frequency of a mass. Unfortunately for Feigl, however, fundamental science has moved on since he was writing. Theories like string theory (or M-theory) and loop quantum gravity, which stand at the forefront of the effort to develop a single 'theory of everything', attempt with increasing success to reduce all such qualitatively distinct properties to a single underlying phenomenon (strings or gravity loops, for example).

While a test has yet to be developed that will clearly prove any one of these advanced theories correct (or its competitors all wrong), the unifying approach is common to them all. As such, it seems to me to be the best option to treat fundamental science as claiming that on a strict analysis, the apparent qualitative distinctions within the physical realm disappear.

If this is correct, of course, it falls upon the idealist to explain how the single quality posited by these theories can turn out to be non-physical, just as it falls upon the physicalist to show how qualitative distinctions can emerge from the qualitatively uniform fundament. The (substance) dualist, on the other hand, is forced to claim that monistic physical theories can never be complete, because they must necessarily omit the mental, which may or may not be ultimately tenable (or comfortable).

So for present purposes, we can dismiss Feigl's claim that there are (fundamental) qualitative differences within the physical. The second half of Feigl's challenge is more

difficult to deal with, but fortunately some of the groundwork has already been laid. Feigl points out that we are able to make apparently-quantitative judgements about some mental phenomena by means of comparisons which require a clear notion of magnitude. For example, if I look out of my window at the sun falling on the houses opposite, I can judge that experience to be brighter than my experience of looking at the wall next to my window[17]. Again, I can judge that both these experiences are brighter than my experience of lying in bed at night with the light off.

The idea here is that there is ultimately nothing more to a quantitative judgement than the placing of its object in a particular place in an ordered series (or indeed, in a particular place in a multidimensional field whose axes are ordered). It's tempting to suggest, against Feigl, that in fact for something to be quantitative in this sense, one also requires a scale of measurement with a 'zero' point, but one reason for using brightness as an example is that there is a (possible) experience of zero brightness - eigenlicht, the 'colour' seen in perfectly sealed, lightless room.

I allow the point that apparently-quantitative ordering is possible of at least some kinds of mental states or qualities. Certainly, one-dimensional qualities (pitches, intensities of light and sound etc.) can be ordered by subjects in a quite accurate way. Similarly, given a particular shade of colour and asked to find its place in a colour-chart, we would expect most if not all subjects to find the correct place.

However, I would argue that these comparative judgements are not *strictly* quantitative, in much the same way that the apparent locations of my experiences are not strictly spatial. It is a matter of the possibility of *exact, systematic* measurement, and each of these requirements presents a problem.

The problem we encountered with the apparent spatiality of experiences 'within my body' was that it was not possible to systematically measure the distance from them (or points within them) to points of physical space. This generalises to any judgement of quantity thus; for a judgement to be properly quantitative, it must allow us to put the quantity at a definite point on a (numerical) scale (or scales) of all quantities of its type[18].

---

[17] Let us assume for now that it is possible either to speak of the aggregate brightness of a complex experience like 'the view from my window' or to break the composite up into elements of which at least some have internally uniform levels of brightness.

[18] To put this more stringently, there must be an (in-principle-measurable) exact relation of the magnitude of the given quality to all other magnitudes of that quality.

This is the problem of systematicity: not only must I be able in principle to judge how a given shade of blue from within my experience relates in magnitude[19] to all other blue-experiences I have had (itself a tenuous enough claim, since it assumes I am capable of exact judgements about the magnitudes in question), but I must be able to relate it in magnitude to all of *your* blue-experiences. Feigl's system of direct comparisons will not allow this in light of our defence of absolute subjectivity; I simply cannot in any way apprehend your blue-experience to be able to draw a comparison between it and mine.

The problem of exactitude has to do with units of measurement, but ultimately boils down to the same key point. A quantitative judgement or measurement must be made in terms of a number (quantity) of units[20], so for my judgements about mental phenomena to be strictly quantitative, a system of units is required.

We should, I think, steer clear of any attempt to use physical units for the mental[21]. To do so would be to cease making purely mental judgements, and to instead make judgements which would turn out to be quantitative only in respect of the physical phenomena they assess. There is nothing wrong with such judgements - they are often very useful on an ordinary, day-to-day basis - but they cannot be what interests us here. This rules out not only the possibility of quantifying the brightness of my visual experiences in terms of the amplitude of an electromagnetic wave, but also the idea that we might quantify it in terms of neural activity.

Feigl does not consider any problem of units, but if he were confronted with it, I would expect his answer to be along these lines: assume that any quantity scale is such that it has a 'zero' value, and measure the magnitude of anything on that scale by counting how many of the smallest measurable steps above zero one must take to reach it. That is, I will find the magnitude of a given quality by counting up the scale from the zero point in the smallest steps that I can distinguish. If I can distinguish a thousand different degrees of brightness, I can then, hypothetically at least, make a quantitative judgement about a given brightness-of-an-experience by counting how many degrees it is from one or other end of the scale.

---

[19] or magnitudes, since we generally place colours in a multidimensional field.

[20] I take this to be part of the definition of 'quantitative'.

[21] This goes even for temporal units. The time we actually experience is not evenly distributed across the time measured by a physical clock, and neither idealists nor physicalists should be allowed to shy away from the task of explaining how this distortion is possible (I do not think this is an insurmountable task in either case, though I will not attempt such a construction in this text).

I have no problem with this idea within its limits. Remember that my stipulation was only that the exact magnitude of a given quantity should be *in principle* specifiable. This means that, while to do so would be tedious to the point of implausibility, I could theoretically compare the experience I wish to quantify to every experience of lesser magnitude that I could distinguish, and thus place the experience in question accurately on its scale.

The problem again comes when we try to make sense of this intersubjectively. It seems to me obvious that it is at least *possible* that there could be a subject whose discriminative abilities in respect of a given experiential quantity significantly differed from mine; let us imagine that within the same range of brightness-stimuli as I can apprehend, he can discern twice as many levels. The units he uses to make quantitative judgements will then be half the size of mine.

This may not seem like a problem, prima facie, since it seems like we can just double all my numbers or halve his to enable us to compare. However, in order to specify the scenario, we had to make use of a physical quantity - the 'range of brightness-stimuli' - and so such comparative judgements will be hybrid and not purely mental.

We can sidestep this if we assume that I and this other, more discerning subject are only capable of experiences in the same range, but there is good reason to reject this assumption. It is entirely possible that he will also be capable of apprehending a broader range of experience-magnitudes.

For example, assume that in total he is able to discern two thousand, one hundred different brightness levels of experiences. The only purely mental data we would then have on which to build an intersubjective unit-set for quantitative assessments would be his claim 'I can discern two thousand, one hundred different brightness levels' and mine, 'I can discern one thousand different brightness levels'. We could standardise these assertions to some set of units if we measured exactly which amplitudes of light waves corresponded to each brightness level that each of us could see, but this is a physical quantity, not a mental one.

Even given statements of this kind from all subjects under consideration, we could not standardise comparisons to one set of units without invoking a physical quantity. Thus, the purely mental offers no solid foundation for a set of intersubjective units in which properly quantitative judgements could be grounded.

The Problem of Judgements

We have thus established that there is a clear separation between qualitative and quantitative judgements which matches the mental/physical distinction; that is, qualitative judgements will always be about the mental, and quantitative about the physical. The problem with drawing the distinction in this way, though, is that it opens us up to accusations of begging the question, because a judgement would ordinarily be thought to be an essentially mental thing.

There is nothing in the use of a judgement criterion of the qualitative/quantitative distinction to prevent a physicalist reducing the mental away; if nothing else, a physicalist can argue that all qualitative judgements are simply vague quantitative judgements, which would (if it turned out to be a plausible step) constitute a reduction of the mental to the physical. This gives the idealist one motivation already for wanting a different criterion of the qualitative, but there is another, more serious problem for idealism: we want all judgements themselves to turn out to be mental, whether they are qualitative or quantitative. That is, the distinction needs to be located in what is judged rather than in the judgement itself.

It's therefore very much in the idealist's interest to find a different way of separating the qualitative from the quantitative. In essence, we need to be able to separate elements of reality in a clear way without having to rely on any mental terms to do so. As mentioned briefly above, I'll adopt the term 'phenomenon' for the most general sense of 'an element of reality'.

I think the best form of the qualitative/quantitative distinction for the idealist is one which uses a distinction between 'real' and 'derived' units. We have already seen that a problem about units foiled Feigl's claim that there were truly quantitative elements of the mental field. The position I am about to outline is a generalisation of the point made above.

Consider this: when I say 'a metre is one hundred centimetres long', we know by definition that those centimetres are identical. Each centimetre is the same length, and they are distributed evenly across the length of the metre. By contrast, if I say 'I can discern a thousand divisions within my visual field'[22], it is impossible to tell if those divisions are

[22] Clumsy as it is, this phrasing is still not quite what I mean. The actual claim is more like 'If I were to move an object across my visual field in the greatest number of steps such that I could discern each one, it would take a thousand such steps'.

evenly distributed, or the same width, without recourse to some other information[23]. This is the essence of what I mean by saying that centimetres are a 'real' unit, while minimum discernible divisions of my visual field are 'derived'.

By way of a clearer illustration, we can look at the pitch of a sound. The physical, 'real' unit of pitch is the Hertz; the number of times a wave - in this case a sound wave - peaks in a second. However, we have a second scale of pitch, the musical scale. The steps (semitones) of the musical scale are clearly 'derived' units; they are logarithmic to the scale of pitch in terms of frequency/Hertz. That is, the step between B and middle C is larger than that between A sharp and B, and smaller than that between middle C and C sharp.

The semitones of the musical scale are not (or at least are not clearly) related to the minimum discernible steps of human pitch-perception, though they do make a good example of a derived unit. More importantly, we know by experiment (Benesty, Sondhi and Huang, 2008. P.65) that the minimum discernible steps of human pitch-perception are not related linearly to frequency; the higher the frequency, the greater the required step will be (in Hertz) before a human will hear a difference.

It's worth noting at this point that there are other ways in which one kind of units can be derivative of another. For example, in apparent contradiction with the example given above, centimetres are officially derivative of the metre - units are always derivative of some standard. Similarly, some quantities are measured in units derivative of the units of other kinds of quantity; speed, for example, is measured in *metres* (distance) per *second* (time).

Units derivative in either of these ways may still turn out to be real, however. A useful way of telling what kind of unit you are using is this: a unit is real if you can ask someone else to measure a given quantity in it and expect them to be able to give a correct answer (or, perhaps even more pertinently, if you can ask someone else to check and confirm your own measurement). If I ask you to measure the width of my desk in centimetres, you can do so with a sufficiently long ruler. If I ask you to measure the same quantity in minimal discernible steps of my visual field, you have no way of knowing which particular step I have in mind, never mind that any such measurement will be relative to how far from the table I am[24].

---

[23] Continuing from the previous footnote, the obvious way to do this would be to use a hypothetical accurate ruler of some description to measure the changes in position of the hypothetical object.

[24] In fact, as shown in our previous discussion of intersubjective units, there would be no sense in the question at all; given absolute subjectivity, you have no way whatsoever of using

This, then, is the key difference between the qualitative and quantitative: if a phenomenon can be measured in real units, it is quantitative. If not, it is qualitative. It might seem that we are no better off than before, because 'measurement' seems to be a concept that presumes awareness, but this misses the point. The problem with our earlier definition was that it was a classification *of judgements*.

That is to say, it divided judgements into two categories, one qualitative and thus about the mental, the other quantitative and thus about the physical, but because judgements are themselves mental, it begged the question between the physicalist and the idealist. It was also, from the idealist's point of view, circular, since it defined (at least partly) the mental in terms of a mental thing.

This new way of differentiating qualitative and quantitative has no such problem, since it draws a global distinction; all phenomena are either measurable in real units or not, which is to say they are either quantitative or qualitative[25]. The question of whether the mental is the qualitative (or merely is qualitative) is entirely separate from this definition, though I think at the very least it is clear that the mental is coextensive with the qualitative.

---

discernible steps of *my* visual field as units, except by first translating them into an intersubjective (and thus non-mental) standard.

[25] To clarify, 'phenomena' here is being used in a fine-grained sense, such that if an entity has both qualitative and quantitative features, these are taken as separate phenomena. A beam of light may have a real-unit-measurable wavelength (a quantitative feature) and a non-real-unit-measurable phenomenal colour (a qualitative feature), but these are separate phenomena.

2. The Refutation of Compatibilism

The Collected Definition of the Mental

We have arrived at a definition of the mental; a thing is mental if it is subjective, non-spatial and qualitative. That is, it is mental if it stands in some relation of privileged access, does not occupy a position in real physical space, and is not measurable in real units. At the moment, this collection of properties seems arbitrary; we have seen only hints of how they relate to one another. The business of this chapter will be to explain why it is that these apparently disparate properties go hand in hand.

In doing so, we shall see that together, the three properties constitute a predicate - 'mental' - whose application is limited to only those things which are not physical, where 'physical' is taken to apply to the collection of phenomena which are non-subjective[26], spatial, and quantitative. We shall also see that every phenomenon must be either mental or physical - that there is no third category. I take these two claims together to constitute a refutation of compatibilism.

In the introduction, I offered the following definition of the four main ontological positions about the mental and physical:

- Dualism: The view that there are irreducibly mental and irreducibly physical entities.
- Physicalism: The view that there are irreducibly physical entities, to which all mental entities in some way reduce.
- Idealism: The view that there are irreducibly mental entities, to which all physical entities in some way reduce.
- Compatibilism: The view that there is a single kind of substance, neither irreducibly mental nor irreducibly physical, in which mental and physical attributes inhere (for some intuitive notion of 'inherence').

We can now refine this somewhat:

- Dualism: The view that there are two distinct kinds of substance, one wholly mental and the other wholly physical.
- Physicalism: The view that there is one kind of substance, and it is wholly physical.

[26] I use the term 'non-subjective' here because in the preceding chapter I allowed that there can be objective facts about the mental/subjective. Tempting as it is to introduce a secondary definition of 'objective', I will refrain in the interests of clarity.

- Idealism: The view that there is one kind of substance, and it is wholly mental.
- Compatibilism: The view that there is one kind of substance, neither wholly mental nor wholly physical.

This refinement is more perspicuous because it shows that compatibilism is more akin to physicalism and idealism than to dualism - it is a monistic theory. Our question, then, should be whether or not a single kind of substance can exhibit both mental and physical behaviours, either simultaneously, as in stronger forms of compatibilism like panpsychism (the view that fundamental particles have both physical and mental properties), or separately, as in Russellian monism (the view that there is only one kind of substance, neither intrinsically mental nor intrinsically physical) and property dualism.

If it seems too ambitious to attempt to refute all forms of compatibilism in one fell swoop, the task of this chapter can be looked at in another way. After all, from a metaphysical point of view, I have very little quibble with many forms of compatibilism. The problem is one of language, and as philosophers we have a degree of freedom in the way we use our terms.

My contention could thus be reconstrued again, as the claim that compatibilist theories all make revisionary use of the terms 'mental' and 'physical', by omitting a key component of the common-use terms: their mutual exclusivity. There is no great rational sin in this revision - with the exception of substance dualism, all these theories are committed to a degree of revision - but I feel it is important to understand that it perhaps fails to reflect the metaphysics of monism, which will become significant for our case in later chapters.

A final word of warning: I have found it immensely useful, in the course of this chapter's argument, to lay things out formally, so what follows will be structured around setting up a series of formal derivations[27]. For the benefit of those less comfortable with formal logic, however, the derivations themselves have been relegated to the appendix; I will attempt to summarise them along the way. In these summaries I shall run roughshod over formal best practice in a way that will bring steam from the ears of the logically astute - whose attention should be instead directed to the appendix.

[27] The formal language and derivation system I am using come from Tomassi's *Logic* (1999).

The Coextensivity of Subjectivity, Qualitativity and Non-spatiality

We have seen in chapter 1 that Feigl's single-property approach to defining the mental is inadequate, in that it fails to provide a complete definition of the mental as something distinct from the physical. Actually, to put it this way is a little unfair; Feigl's argument is of the form 'no single property adequately defines the mental, so the mental cannot be adequately defined'. This, of course, requires the implicit premise that an adequate definition of the mental can only be given in terms of a single property. Is this premise justified?

The answer certainly cannot be an outright affirmation. The vast majority of things in the world are defined in terms of multiple properties. No single property defines a human being, for example, or a piano. However, there is something to be said for Feigl's implicit premise when it comes specifically to the question of the mental. This is because it seems wrong to identify the mental with the intersection of arbitrarily-selected properties. It would lead to the distinct oddness of there being, at least potentially, entities which were 'partly' or 'half-way' mental.

So, if there are non-spatial things which are not also qualitative and subjective, for example (and Platonic Forms might be considered an example of this), it would be difficult to engage with the question of whether or not they were mental phenomena. If nothing else, this would leave the door wide open for compatibilists.

On the other hand, if it could be shown that subjectivity, qualitativity and non-spatiality were necessarily coextensive, it would be strong evidence that there is some special category of phenomena for which it would be (as I have claimed) natural to reserve the word 'mental'. And I believe just such a demonstration is possible.

So, setting our domain to 'all existents', let:

P = '... is subjective'
Q = '... is qualitative'
R = '... is spatial'

We are trying to show that these three qualities are coextensive (or rather, that subjectivity and qualitativity are each coextensive with non-spatiality). The simplest way to state this formally is '$\forall x[(Px \leftrightarrow Qx) \& (Px \leftrightarrow \sim Rx)]$'[28] (which of course implies '$\forall x[Qx \leftrightarrow$

[28] One clarification that is needed here is the modal issue. In order to save having to mix modal and quantificational logic, I am treating these logical formulations as expressing

~Rx]'); informally, we can read this as 'anything that is subjective is qualitative, and vice versa, and anything that is subjective is not spatial, and vice versa' (which implies by the transitivity of material implication that anything which is qualitative is non-spatial, and vice versa).

To prove this, we need four premises:

1. ∀x[Px ⇾ Qx] (anything that is subjective is qualitative)
2. ∀x[~Px ⇾ ~Qx] (anything that is not subjective is not qualitative)
3. ∀x[Px ⇾ ~Rx] (anything that is subjective is non-spatial)
4. ∀x[~Qx ⇾ Rx] (anything that is not qualitative is spatial)

With some hand-waving to cover for quantification, modus tollens on premise 2 will give us ∀x[~~Qx ⇾ ~~Px] and thus ∀x[Qx ⇾ Px], which, with premise 1, gives us ∀x[(Px ⇿ Qx)]. Modus tollens on premise 3 will give us ∀x[~Rx ⇾ ~~Qx], and with the previous sentence this gets us to ∀x[~Rx ⇾ Px], and thus to ∀x[(Px ⇿ ~Rx)]. A final act of conjunction will give us the target formula. The full derivation can be found in the appendix to this chapter.

It remains only to be shown that the premises we have used are true. Taking them one at a time:

Anything that is subjective is qualitative.

Quantitativity requires measurability in real units; a phenomenon is only quantitative, in the way I have defined the term, if it is measurable in terms of something which can be counted by multiple different people. But all subjective phenomena are privileged. It is not possible to compare them to anything except other phenomena privileged to the same individual subject. Anything which can objectively be known about them (and as we saw earlier, the idealist is going to have to acknowledge that the *existence* of privileged contents is a matter of objective fact even if their contents are privileged) cannot fully capture their qualities.

It may be that some standards by which a given subject can evaluate his or her own mental states behave in ways similar to real units. For example, it may be that across the central part of my visual field, my ability to discern movement is equally fine-grained, so that any of a number of discernible steps may be interchangeable with another of that number in the way that any centimetre is interchangeable with another. This would not

---

conceptual/analytic truths - that is to say, each claim should be taken as being made for all possible worlds.

make those steps real units, however; not only are they part of a sequence which may also include units which are derived, in the sense I have defined, but I would remain the only person who could use them.

I said above that a key part of what it is for a unit to be real is that measurements using it can be checked by others; no such checking is possible in the case of purely subjective units, so they cannot be real. Thus, the privileged portion of any subjective phenomenon is not going to be measurable in a real unit, because no real unit could be privileged to a particular subject. And thus, it cannot be quantitative.

Anything that is not subjective is quantitative.

Given the preceding, this may sound like it should be obvious, but it's not quite that simple. A thing's not being subjective only means that there will be some intersubjective standard to which it is comparable. That standard doesn't, prima facie, have to be a real unit, which is what would be required for quantitativity.

In fact, one could raise a question over whether a thing's being objective actually implies it must be comparable to some other thing, since it is at least conceivable that there could be an intersubjective phenomenon which was only perceivable in a unique mode which produced experiences not comparable to any other. However, short of invoking a divine mystery, or something of that ilk, it is hard to see what could count as even a hypothetical example of this.

The more important question is whether all intersubjective standards for comparison must be real units. If they are, then the non-subjectivity of a thing will be sufficient to imply its quantitativity, but it is not obvious that they are. We might naturally think that plenty of ways we can compare things make no mention of any units at all; for example, if you are telling me about a particular occasion on which you saw an unusually brightly blue sky, I might ask you whether the sky you saw was a brighter blue than the sky currently above us.

Prima facie, there are no real units involved in your answer, whether it's a 'yes' or a 'no'. But when we look harder at what you *mean* by saying the sky now is less or more brightly blue than the one you saw, I think it is quite hard to make sense of your utterance without ultimately invoking some real unit.

Certainly, when you in your own head relate your experience of the blue of the current sky to that of the remembered sky, no real unit is needed. If units are needed at all, minimum perceivable steps in your own visual sensitivity will do just fine, and as discussed in chapter one, these are not real units in the relevant sense. If the sky today is not as

bright as the one you are reporting about to me, then when you say so, the meaning of that utterance in your head can be fully captured without the inclusion of any physical quantity.

But, assuming that I do understand your answer and come to (correctly) understand that the sky to which you have been referring is brighter than the one I can see by looking up, we have involved a physical phenomenon, and one which is not privileged to either of us - namely, the actual physical properties of the light currently being bent through the atmosphere towards us.

Generally speaking, it would be trivial to point out that our ability to compare our experiences with each other's depends on a shared frame of reference, but here the precise nature of that frame becomes deeply significant. We must make two assumptions: first, that our perceptual systems are significantly similar, and second, that your eyes and mine are actually currently receiving light of significantly similar colour and brightness. Without this baseline to work from, you telling me that the sky you remember is brighter than the one you are looking at right now would be no more informative than if we were conversing by phone from different continents, looking up at completely different skies.

And we cannot treat the brightness[29] of the shared sky as anything other than a quantity of real units. Remember, we're only talking about things which we already know to be non-subjective (i.e. non-privileged). The phenomenal contents of our separate experiences of the sky are privileged, and thus off the table. Only the actual physical quantities making up the light our eyes absorb are shared.

Anything that is subjective is non-spatial.

We have already seen that if a thing is subjective, it must be qualitative (and thus not quantitative). But to be spatial, a thing must have exactly-specifiable distance relations, not just to every other spatial thing but also to and between its parts (more on parthood and spatiality later), and distances are paradigmatically quantitative. The spatial dimensions of

[29] To be completely precise, the 'bright-blueness' of the sky is a composite phenomenon involving various real units - the frequencies and amplitudes of all the light waves making up the signal entering my eye - and even if we tried to simplify by using a system similar to the way print inks or computer screen pixels are combined to produce colours, we would still end up with a multidimensional quantity (variations on the theme of 'red: 0, green: 0, blue: 255', or something similar). For present purposes, then, let us assume that there is some way of crunching all these numbers to produce a single scale of 'bright-blueness'. Despite being derivative of all sorts of other quantities, this scale will still be a real-unit scale, because its units can be evenly distributed along its length, and will be usable by anyone who can follow the maths.

a thing must be measurable in real units[30]. Thus, if a thing is spatial, it must be quantitative, and thus not qualitative, and thus not subjective.

Technically, what this argument shows is actually ∀x[Rx ⇾ ~Qx] (anything that is spatial is not qualitative). Given what we've already discussed, however, we know that ∀x[Px ⇾ Qx], and thus by modus tollens, ∀x[~Qx ⇾ ~Px]. So ∀x[Rx ⇾ ~Px], and one more round of modus tollens, plus eliminating a double negative, will give ∀x[Px ⇾ ~Rx]. The latter form was chosen to keep the main derivation simpler.

Anything that is not qualitative is spatial.

I've left this premise to last because it is far and away the hardest to defend, and to do so I will want to rely on preceding premises. The problem is not a technical matter - it is simple (as I shall show) to construe any quantity in a spatial way - but a matter of making sure that the quantitative turns out to be spatial *in the right way*.

Most physical phenomena are the products of several different quantities. For example, the colour displayed at a point on a computer monitor is a product of three (depending on the system used, the set will either be red, green and blue or hue, saturation and luminescence). Each shade has a unique combination of these three values. So we could define an abstract three-dimensional space which had axes of red, green and blue or hue, saturation and luminescence, and every colour on the monitor screen can be mapped to some point in this abstract space.

In this way, then, any quantity can be placed in some abstract space or other, even if that space only has a single dimension. This, however, does not suffice for our purposes. It was a key point in our discussion of spatiality that the spatiality we are interested in is *genuine* spatiality - a location in the real spatial field that our physical bodies inhabit. And the fact that all quantities are located in what we might call 'quantity spaces' does nothing to guarantee this.

Carrying on with our monitor example, the point is that knowing where the colour of a given pixel sits in the abstract colour-space does not tell us if it will appear on the screen at all, or when or where. What is needed is some way of guaranteeing that if a thing is quantitatively measurable, it must be present in real space.

---

[30] Care is needed here to deal with the elasticity and curvature of space proposed by the general theory of relativity, but to go into the level of detail required would be beyond the scope of this work. Suffice it to say for now that I am confident that either it will be possible to deliver a set of real-unit axes for the space described by general relativity, or the universe will turn out to be fundamentally non-spatial (because wholly mental).

The best strategy I have found for arguing that all quantities must appear in a real physical space is to insist on a strict understanding of 'measurable'. We saw earlier that the key feature of things that are quantitative – measurable in real units – is that any measurement of them is checkable by multiple subjects. We can take 'measurable' to mean that there is some conceivable act of measurement of the phenomenon in question.

The final step would be to insist that all acts of measurement require it to be possible for the object being measured to be spatially near a measuring object. That is not to say that the object need exist in a universe containing an actual measuring object, only that it exist in a universe *capable* of containing a measuring object. To insist that such a universe be spatial is, I think, not such a great stretch from there, since this is how we most naturally understand relations between objects – to suggest that an act of measurement could take place non-spatially would be the more outlandish position, in need of substantial defence.

Are there candidates for an exception to this requirement? Perhaps the most obvious would be a quantity in abstract; the universal, for example, of some specific shade of blue[31]. However, I am not satisfied that such a universal is ever itself measured – instead, I would say that the universal is what a certain phenomenon, its instance, *is measured to be*. The universal itself is non-quantitative and mental (I am aware this is a position of some controversy, but it is a controversy too large to address), whereas its instance, if it is to be measured, must be capable of being brought into spatial proximity with an appropriate measuring implement.

I find this suggestion very plausible, though I'm not sure I can see a way to make it rigorous. A point in its favour is the dearth of obvious alternatives. The idea of a measurable quantity for which there is no conceivable act of measurement seems to me to require a substantial imaginative leap towards very hand-wavy, obscure scenarios. I am happy to count on that fact for now.

The Mutual Exclusivity and Exhaustivity of the Mental and the Physical

What all this shows, in essence, is that subjectivity, qualitativity and non-spatiality are coextensive; so the same goes for non-subjectivity, quantitativity and spatiality. Thus if we choose to define the mental as subjective, it must also be qualitative and non-spatial.

[31] If, indeed, such a thing exists. As a nominalist, I am no friend to the idea of universals in general, and even less to the idea of universals of specific colour shades, except as intermediary derivatives of the more general colour universals. This, though, is much too vast a debate to get into here.

Likewise, if we choose to define the physical as spatial, it must also be objective and quantitative.

Continuing with our interpretation from before, let's introduce a two new predicates:

S = '... is mental'
T = '... is physical'

We can formulate the definition of the mental as ∀x[Sx ⇿ ((Px & Qx) & ~Rx)] and of the physical as ∀x[Tx ⇿ ((~Px & ~Qx) & Rx)]. What we are trying to show, ultimately, is not merely that the mental and the physical are mutually exclusive (∀x[~(Sx & Tx)]), but also that they are *exhaustive*; that everything is either mental or physical; ∀x[Sx v Tx].

Given ∀x[Sx ⇿ ((Px & Qx) & ~Rx)], it should be obvious that ∀x[Sx ⇾ ~Rx] (Px or Qx could just as easily replace the ~Rx; the selection is purely arbitrary). By the same logic, ∀x[Tx ⇿ ((~Px & ~Qx) & Rx)] gives ∀x[Tx ⇾ Rx]. So for any object a, (Sa & Ta) will imply (Ra & ~Ra), which suffices to show that ∀x[~(Sx & Tx)].

Going the other way requires rather more work. The key is to remember that ∀x[Sx v Tx] is equivalent to ∀x[~(~Sx & ~Tx)] - what matters is what would happen if, for some object a, (~Sa & ~Ta). By ∀x[Sx ⇿ ((Px & Qx) & ~Rx)], we know that ∀x[((Px & Qx) & ~Rx) ⇾ Sx], and thus by modus tollens ∀x[~Sx ⇾ ~((Px & Qx) & ~Rx)]. The same steps can be performed on ∀x[Tx ⇿ ((~Px & ~Qx) & Rx)] to give ∀x[~Tx ⇾ ~((~Px & ~Qx) & Rx)].

We need to show, then, that for any object a, ~((Pa & Qa) & ~Ra) and ~((~Pa & ~Qa) & Ra) are contradictory. For this we fall back on the conclusion of the previous section, ∀x[(Px ⇿ Qx) & (Px ⇿ ~Rx)]. ~((Pa & Qa) & ~Rx) is equivalent to ~(Pa & Qa) v Rx, which is equivalent to the disjunction of ~Pa, ~Qa and Ra. By ∀x[(Px ⇿ Qx) & (Px ⇿ ~Rx)], though, we know that ~Qa ⇾ ~Pa, and that Ra ⇾ ~Pa; the three disjuncts will all share any consequence of the first.

~Pa implies both ~Qa and Ra (again, by ∀x[(Px ⇿ Qx) & (Px ⇿ ~Rx)]), so any time that we have ~Pa, we can get (~Pa & ~Qa) & Ra. So ~Sa will imply (~Pa & ~Qa) & Ra; but we know from ∀x[~Tx ⇾ ~((~Px & ~Qx) & Rx)] that ~Ta will give ~((~Pa & ~Qa) & Ra). For both ~Sa and ~Ta to be true would be contradictory. Therefore, ∀x[~(~Sx & ~Tx)], i.e. ∀x[Sx v Tx].

The Refutation of Compatibilism

Showing that a thing cannot be both mental and physical deals with stronger forms of compatibilism, such as panpsychism. It might be thought that showing that everything is

either physical or mental rules out weaker, neutral-monist forms, since neutral monism presumably by definition requires that there be a substance which is neither mental nor physical.

It may be possible, however, to draw up a plausible argument that there is a single kind of substance, parts[32] of which are mental and other parts physical. This would be a very weak form of compatibilism indeed (and in fact, I think many would regard it simply as substance-dualism-by-another-name), but it may still be a viable position. It therefore will be worth our while, if possible, to refute it.

To do so, we will need to extend our interpretation yet further. At least, we will need to introduce a relation of parthood:

G = '... is a part of ...' (i.e. 'Gxy' means 'x is a part of y')

We can thus state the conclusion we seek as $\forall x[\forall y[\forall z[(Gxz \& Gyz) \nrightarrow \sim(Sx \& Ty)]]]$, which can be read as 'if x and y are both parts of z, then it cannot be the case that x is mental and y physical'. This amounts to a refutation of this final, weakest form of compatibilism.

To prove this conclusion from what we've already established, however, we will need some more logical apparatus. First, two more predicate terms:

F = '... belongs to the field ...' (i.e. 'Fxy' means 'y is a field to which x belongs')
U = '... is a (real, genuinely) spatial field'

These are awkward terms, logically (we should perhaps have a separate predicate for 'being a field', but this would just add clutter to the derivation). For present purposes we can rely on a roughly intuitive notion of what counts as a 'genuinely spatial field' - as hinted at previously, a genuinely spatial field is one which is structurally identical to the one we actually inhabit. More clarity can perhaps be introduced by a discussion of the premises we must introduce in order to reach our conclusion.

The first of these is a formal definition of our earlier predicate of spatiality, R:

$\forall x[Rx \Longleftrightarrow \exists y[Fxy \& Uy]]$

---

[32] This may be a slightly unusual notion of parthood. What we are arguing against is the possibility of an object having both mental and physical aspects, and treating these as conventional, compositional parts is a stretch. However, I do not think that the stretch required is an untenable one.

This means that a thing is spatial if and only if there is some genuinely spatial field to which it belongs. Assuming that there is a legitimate meaning to being a 'genuinely spatial field', I don't think there's anything particularly controversial to this premise[33], particularly if we add the stipulation from earlier that to belong to a spatial field, a thing must have specifiable distance relations to every other member of that field.

So, the genuine spatiality of a field is what makes that field sufficient that membership of or belonging to it renders a thing spatial. This doesn't add much by itself. The second new premise that we require, however, should hopefully bring out the most important consequence of being a genuinely spatial field:

$\forall x[\forall y[\forall z[(Uz \ \& \ Gxy) \rightarrow (Fyz \leftrightarrow Fxz)]]]$

This says that if x is part of y and either x or y belongs to a genuinely spatial field, the other must too. It is a modification of a much simpler formula relating part-hood and field-membership:

$\forall x[\forall x[\forall z[(Gxy \ \& \ Fyz) \rightarrow Fxz]]]$

i.e. 'If x is part of y and y belongs to the field z, x must belong to z'. This, I think, is uncontroversial. However, our proof will require being able to make the jump in the opposite direction, too; to show that if any *part* of a thing belongs to a (genuinely spatial) field, the whole thing must as well. This is certainly not true for all spatial and quasi-spatial fields; to take a literal example, by reaching through a fence, I can put my arms into a field on the other side without the rest of me belonging to it. Similarly, an audiovisual phenomenon such as a film or TV program has a visual part which must belong to a visual field, where its audio part will not.

This is where the predicate U begins to do its real work. We can perhaps add to the definition of 'genuinely spatial' an exhaustiveness clause; that is, a field is only *genuinely* spatial if there are no spatial (distance) relations linking it to anything that isn't an occupant of it. Thus the field into which I insert my arms through gaps in a fence is seen as a subset of the complete spatial field of our universe[34].

---

[33] Again, it will need minor adjustment for a Leibnizian framework - 'belonging to' reconstrued as 'constitutively sustaining' - but I don't think this will have any effect at the formal level.

[34] This will go for any n-dimensional space embedded in a higher-dimensional space, too. So for example, if we were interested in a particular 2-dimensional space that was contained within a 3-dimensional space, the 2-dimensional space is only genuinely spatial in virtue of being a sub-space of

And I think that, construed this way, it becomes clearer that my arms being spatial is sufficient for me to be spatial as well. Technically what we have arrived at is:

$\forall x[\forall y[\forall z[Uz \rightarrow (Gxy \rightarrow (Fyz \leftrightarrow Fxz))]$

But this clearly implies the premise stated above, and the earlier version cuts a few steps out of what is already a very long derivation.

What remains, then, is the question of whether this rather stringent premise also applies to parts in the particular, relevant sense; that is, the sense in which an object (such as a person) may have both a mental and a physical part. To assert that a person is a composite of a mental object (a mind) and a physical (a body), is classical Cartesian dualism, and nothing I consider a threat; to assert that a person is one single object with mental and physical aspects is the last form of compatibilism remaining to be refuted[35].

In this form of compatibilism, the mental and physical, though distinct, must be bound together more tightly than the two sides of a piece of paper. They would be analogous to the two sides of a piece of flat material of absolutely minimal thinness, so that no process could separate them; not composed of distinct lumps of stuff at all, but different sides of the same stuff. The one side could not be in a real physical space without the other, since they are from a material point of view one and the same thing.

As ever the full derivation is in the appendix, but here is the summary:

Since the target formula is a conditional, we start by assuming the antecedent (or rather, since the target is a universally quantified conditional, we're assuming an instance of the antecedent): (Gac & Gbc). This simply picks out three arbitrary objects, a, b and c, such that a and b are both parts of object c. It's trivially true that c is either an object in space or not, so we can freely introduce (Rc v ~Rc). This gives us a starting point for a derivation whose main structure is disjunction-elimination; showing that the desired consequent, ~(Sa & Tb) (or 'it is not the case that a is mental and b physical') follows both from Rc and ~Rc.

Starting from Rc - c's being spatial - first, we know that if c is spatial, it must belong to a genuinely spatial field, which we'll call d, so (Fcd & Ud). We can now show that a

---

the (presumed exhaustive) 3-dimensional space. I am also taking it for granted that two partially-overlapping spaces must by definition be sub-spaces of a space which at least encompasses them both (but which need not be ontologically prior to them). The details of this are not important for current purposes.

[35] If, indeed, this position can be clearly distinguished from the substance-dualist position just outlined. I am not at all convinced that it can be, but I am not sure I can close off the question at that point.

contradiction results from (Sa & Tb). If Sa is true, then by the definition of 'mental' (S), it must be the case that ~Ra; a is non-spatial. If a is non-spatial, then by the definition of spatiality, it cannot belong to a genuinely spatial field (∀y[~(Fay & Uy)]), and thus it cannot belong to d.

So we know that ~Fad (given that we have stipulated Ud). On the other hand, though, a is part of c (Gac), and we know that C belongs to the field d (Fcd). From the definition of a genuinely spatial field, ∀x[∀y[∀z[(Uz & Gxy) ⇾ (Fyz ⇿ Fxz)]]], we can see that (Ud & Gac) ⇾ (Fcd ⇿ Fad). This implies that Fcd ⇾ Fad, and thus Fad, which gives us the contradiction we're after. This refutes the assumption of (Sa & Tb), and shows that Rc ⇾ ~(Sa & Tb).

For the case in which c is a non-spatial object, ~Rc, we begin in a similar fashion, by assuming (Sa & Tb) and looking for a contradiction. From the definition of spatiality, we know that ~Rc ⇾ ~∃y[Fcy & Uy] by modus tollens. What we'll now show is that Fb ⇾ ∃y[Fcy & Uy].

By the definition of 'physical' (T), we know that Tb gives Rb. From the definition of spatiality (R), we know that Rb gives (Fbe & Ue), where e is an arbitrarily-chosen field. But again, the definition of U, with the initial assumption of Gbc, means that Fbe ⇿ Fce, which gives us Fce & Ue. This can be existentially quantified to produce ∃y[Fcy & Uy], again giving us the contradiction we seek.

So the truth of (Gac & Gbc) entails that (Sa & Tb) will always produce a contradiction, whether c is spatial or not. So there can be no single lump of stuff, be it an individual human body or the universe as a whole, which has both mental and physical parts. This puts paid to the final, weakest form of compatibilism introduced at the start of this section.

Actually, we have to be a little careful with how we understand the claim that no single lump of stuff can have mental and physical parts, since put that bluntly it seems far too strong. It's not intended to rule out substance dualism (which in these terms would be saying that an individual human is two distinct lumps of (different kinds of) stuff, a body and a mind, which are somehow tied together[36]), though it may strengthen the interaction problem faced by classical dualism.

It's also not intended to rule out idealism and physicalism. It forces us, though, to construe idealism as claiming that some parts of the mental realm *simulate* physicality, and to construe physicalism as claiming that some parts of the physical realm *simulate*

---

[36] Obviously, there are far more complex metaphysics at work here than this brief statement can capture, but this is not the time or place to go into them.

mentality. I don't think this is terribly costly. All that is ruled out is the idea that any single lump of stuff could have irreducibly mental and physical parts.

We have thus shown that not only is everything either mental or physical and never both, but also that no single kind of stuff can have both mental and physical parts. I take this to constitute a full and thorough refutation of all forms of compatibilism.

A Sketch of Reductive Strategies

We have defined the mental as the non-spatial, qualitative and subjective. The physical can thus be defined as the spatial, quantitative and non-subjective. It is arguably possible for something to be neither mental nor physical on this classification, but it is not possible for anything to be both. Nor is it possible for a single kind of substance to support both irreducible mentality and irreducible physicality, since any single kind of substance must be either spatial or non-spatial.

Thus, all forms of the compatibilist strategy fail, or at least turn out to be ultimately not compatibilistic in nature. The only ontological positions are physicalism, idealism and dualism. Each position has different work to do if it is to be established as correct.

The physicalist must explain the emergence of the non-spatial, qualitative and subjective from facts about the way we perceive the world. The idealist must show how phenomena which are fundamentally private (and thus not strictly spatial or quantitative) can give rise to the appearance of a public (objective) world. The dualist faces a series of familiar problems to do with the possibility of interaction between separate kinds of substance.

Showing that these arguments are possible does not, of course, imply that the views that require them are true. As such, I will not be spending much of the rest of this thesis laying out how an idealist reduction could function. However, if the physicalist's reduction or the idealist's reduction turned out to be impossible, physicalism or idealism would be false.

This, too, is not the argumentative strategy I plan to pursue, but it is necessary, I think, to show that our definitions of 'mental' and 'physical' do not beg the question in either way. That is, we ought if possible to show that both physicalist and idealist reductions are plausible, so as to avoid the accusation of sneaking in one or the other view by the back door.

For physicalism, we would need to be able to show how the appearance of privilege, qualitativity and non-spatiality which the mental possesses could arise as an illusion. This, I

think, can be hung on the nature of the qualitative. The physicalist is going to want to say that everything that exists (at the fundamental level) is quantitatively measurable. However, he is not committed to the claim that everything that exists is *non-invasively* or *non-destructively* measurable.

A crude example is provided by Schrödinger's famous cat; one can measure the state of the cat only by compromising the experiment and opening the box. So it is, the physicalist can argue, with our minds - we could measure the properties of the mind, but only by sticking instruments into the brain which would cause a subsequent cessation of brain function. The appearance of qualitativity arises only when the brain synthesises (as it does, all the time) vast ranges of quantitative stimuli (e.g. the individual nerve signals produced by each of the millions of photoreceptive cells in the eye) into phenomenal contents that we can actually make use of (the colours we see as a result).

This, too, explains the appearance of subjectivity and non-spatiality. No other subject can access the products of this neurological process because it is not their brain which has produced them. Again, such access could be had to any given mental state, but only by dismantling the mind that held it. As for the appearance of non-spatiality, as we saw in chapter 1, the human nervous system has a limited resolution and so cannot produce a phenomenal field detailed enough to map perfectly onto physical space.

Obviously, this is the briefest possible summary, and whole libraries could be written on how a physicalist reduction of the mental could work, but I hope it shows that I haven't ruled physicalism out by definition. I won't try to show how a dualist could overcome the causal problem of interaction, but I don't think that the definitions of 'mental' and 'physical' that I've offered here has raised the bar for that task any higher than it has always been.

What remains, then, before we move on, is to quickly sketch how the appearance of quantitativity, spatiality and non-subjectivity can arise from the mental as we've defined it. Here again, the qualitative/quantitative distinction does a lot of the work. In chapter 1, we saw that the qualitative realm is only measurable in 'derived' units, such as the minimal discernible steps of a given subject's visual or auditory field.

Intersubjective, 'real' units may therefore be argued to be a product of using some shared object to translate between different subjects' sets of derived units. For example, you and I might be able to see vastly different numbers of minimal discernible steps along the edge of a ruler, but as long as we can both see the ruler, we can compare distances measured with it. In this way, we can account for the arising of the appearance, generally, of quantitativity, and specifically of precise spatial distances.

The big question for idealists, and one far too large to tackle here, is how intersubjective phenomena can arise from the purely subjective. That is, given that everything that exists is privileged to some subject or other[37], how is it that we can share experiences at all? Of course, things only need to stand in a single privileged relation to be mental, and may stand in many (for example, causal) non-privileged relations to other things besides their privileged subject. But assembling a complete world out of such non-privileged relations is a vast task, far too grand for this project.

I don't believe, however, that the scale of that task should be counted as a point in favour of physicalism. It's true that physicalism probably has the smaller task in explaining how the mental comes from the physical rather than vice versa, but physicalism has other problems, problems we shall come to in great detail as we go on.

Before we do that, however, I want to look at the ideas of two idealists whose work fits rather neatly within this schema; Kant and Bradley. Kant's transcendental and Bradley's absolute idealist systems are interesting because their strategies are focussed on the spatial issue - they attempt to show that reality cannot be spatial (we can perhaps hope to extend from this to subjectivity and qualitativity). I have already mentioned Kant's ideas earlier in this chapter; to some extent Bradley can be seen as taking those ideas to their logical extreme.

Appendix: Formal Proofs

The Coextensivity of Subjectivity, Qualitativity and Non-spatiality

$\forall x[Px \rightarrow Qx], \forall x[\sim Px \rightarrow \sim Qx], \forall x[\sim Qx \rightarrow Rx], \forall x[Px \rightarrow \sim Rx] : \forall x[(Px \leftrightarrow Qx) \ \& \ (Px \leftrightarrow \sim Rx)]$

{1}	1. $\forall x[Px \rightarrow Qx]$	P
{2}	2. $\forall x[\sim Px \rightarrow \sim Qx]$	P
{3}	3. $\forall x[\sim Qx \rightarrow Rx]$	P
{4}	4. $\forall x[Px \rightarrow \sim Rx]$	P
{5}	5. Qa	A for CP
{5}	6. $\sim\sim Qa$	5, DNI
{2}	7. $\sim Pa \rightarrow \sim Qa$	2, UE
{2, 5}	8. $\sim\sim Pa$	6, 7, MT

[37] And if we were going into this in more detail, we'd have to discuss at some length the nature of subjects and how they come to be included in the definition of the mental.

{2, 5}	9. Pa	8, DNE
{2}	10.Qa ⇾ Pa	5, 9, CP
{1}	11.Pa ⇾ Qa	1, UE
{1, 2}	12.Pa ⇿ Qa	10, 11, ⇿I
{13}	13.~Ra	A for CP
{3}	14.~Qa ⇾ Ra	3, UE
{3, 13}	15.~~Qa	13, 14, MT
{2, 3, 13}	16.~~Pa	15, 7, MT
{2, 3, 13}	17.Pa	16, DNE
{2, 3}	18.~Ra ⇾ Pa	13, 17, CP
{4}	19.Pa ⇾ ~Ra	4, UE
{2, 3, 4}	20.Pa ⇿ ~Ra	18, 19, ⇿I
{1, 2, 3, 4}	21.(Pa ⇿ Qa) & (Pa ⇿ ~Ra)	12, 20, &I
{1, 2, 3, 4}	22.∀x[(Px ⇿ Qx) & (Px ⇿ ~Rx)]	21, UI

The Mutual Exclusivity and Exhaustivity of the Mental and the Physical

The original target formula in this section was ∀x[~(Sx & Tx) & (Sx v Tx)]. However, the precise mechanics of proving exactly this are extremely laborious, and this formula can be shown to be a consequence of the somewhat easier-reached ∀x[Sx ⇿ ~Tx]:

∀x[Sx ⇿ ~Tx] : ∀x[~(Sx & Tx) & (Sx v Tx)]

{1}	1. ∀x[Sx ⇿ ~Tx]	P
{-}	2. ∀x[Sx v ~Sx]	T[38]
{3}	3. Sa & Ta	A for RAA
{1}	4. Sa ⇿ ~Ta	1, UE
{1}	5. (Sa ⇾ ~Ta) & (~Ta ⇾ Sa)	4, ⇿E
{1}	6. Sa ⇾ ~Ta	5, &E
{3}	7. Sa	3, &E
{1, 3}	8. ~Ta	6, 7, MP
{3}	9. Ta	3, &E
{1, 3}	10.Ta & ~Ta	8, 9, &I
{1}	11.~(Sa &Ta)	3, 10, RAA

[38] I have introduced this as a theorem rather than a premise simply to save time; I trust there should be no doubt of its correctness for a complete system of formal logic such as Tomassi's.

{-}	12.Sa v ~Sa	2, UE
{13}	13.Sa	A for vE
{14}	14.~(Sa v Ta)	A for RAA
{13}	15.Sa v Ta	13, vI
{13, 14}	16.(Sa v Ta) & ~(Sa v Ta)	13, 14, &I
{13}	17.~~(Sa v Ta)	14, 16, RAA
{13}	18.Sa v Ta	17, DNE
{19}	19.~Sa	A for vE
{20}	20.~(Sa v Ta)	A for RAA
{1}	21.~Ta ⇾ Sa	5, &E
{1, 19}	22.~~Ta	19, 21, MT
{1, 19}	23.Ta	22, DNE
{1, 19}	24.Sa v Ta	23, vI
{1, 19, 20}	25.(Sa v Ta) & ~(Sa v Ta)	20, 24, &I
{1, 19}	26.~~(Sa v Ta)	20, 25, RAA
{1, 19}	27.Sa v Ta	26, DNE
{1}	28.Sa v Ta	12, 13, 18, 19, 27, vE
{1}	29.~(Sa & Ta) & (Sa v Ta)	11, 28, &I
{1}	30.∀x[~(Sx & Tx) & (Sx v Tx)]	29, UI

Here, then, is the proof of ∀x[Sx ⇿ ~Tx]:

∀x[Sx ⇿ ((Px & Qx) & ~Rx)], ∀x[Tx ⇿ ((~Px & ~Qx) & Rx)], ∀x[(Px ⇿ Qx) & (Px ⇿ ~Rx)] : ∀x[Sx ⇿ ~Tx]

{1}	1. ∀x[Sx ⇿ ((Px & Qx) & ~Rx)]	P
{2}	2. ∀x[Tx ⇿ ((~Px & ~Qx) & Rx)]	P
{3}	3. ∀x[(Px ⇿ Qx) & (Px ⇿ ~Rx)]	P
{4}	4. Sa	A for CP
{5}	5. Ta	A for RAA
{1}	6. Sa ⇿ ((Pa & Qa) & ~Ra)	1, UE
{1}	7. (Sa ⇾ ((Pa & Qa) & ~Ra)) & (((Pa & Qa) & ~Ra) ⇾ Sa)	7, ⇿E
{1}	8. Sa ⇾ ((Pa & Qa) & ~Ra)	7, &E
{1, 4}	9. (Pa & Qa) & ~Ra	4, 8, MP
{1, 4}	10.~Ra	9, &E
{2}	11.Ta ⇿ ((~Pa & ~Qa) & Ra)	2, UE

{2}	12.(Ta ⇾ ((~Pa & ~Qa) & Ra)) & (((~Pa & ~Qa) & Ra) ⇾ Ta)	11, ⇿E
{2}	13.Ta ⇾ ((~Pa & ~Qa) & Ra)	12, &E
{2, 5}	14.(~Pa & ~Qa) & Ra	5, 13, MP
{2, 5}	15.Ra	14, &E
{1, 2, 4, 5}	16.Ra & ~Ra	10, 15, &I
{1, 2, 4}	17.~Ta	5, 16, RAA
{1, 2}	18.Sa ⇾ ~Ta	4, 17, CP
{19}	19.~Ta	A for CP
{20}	20.~Sa	A for RAA
{2}	21.((~Pa & ~Qa) & Ra) ⇾ Ta	12, &E
{1}	22.((Pa & Qa) & ~Ra) ⇾ Sa	7, &E
{2, 19}	23.~((~Pa & ~Qa) & Ra)	19, 21, MT
{1, 20}	24.~((Pa & Qa) & ~Ra)	20, 22, MT
{3}	25.(Pa ⇿ Qa) & (Pa ⇿ ~Ra)	3, UE
{3}	26.Pa ⇿ Qa	25, &E
{3}	27.(Pa ⇾ Qa) & (Qa ⇾ Pa)	26, ⇿E
{3}	28.Pa ⇾ Qa	27, &E
{3}	29.Qa ⇾ Pa	27, &E
{3}	30.Pa ⇿ ~Ra	25, &E
{3}	31.(Pa ⇾ ~Ra) & (~Ra ⇾ Pa)	30, ⇿E
{3}	32.Pa ⇾ ~Ra	31, &E
{3}	33.~Ra ⇾ Pa	31, &E
{34}	34.Pa	A for RAA
{3, 34}	35.Qa	28, 34, MP
{3, 34}	36.~Ra	32, 34, MP
{3, 34}	37.Pa & Qa	34, 35, &I
{3, 34}	38.(Pa & Qa) & ~Ra	36, 37, &I
{1, 3, 20, 34}	39.((Pa & Qa) & ~Ra) & ~((Pa & Qa) & ~Ra)	24, 38, &I
{1, 3, 20}	40.~Pa	34, 39, RAA
{41}	41.~Pa	A for RAA
{3, 41}	42.~Qa	29, 41, MT
{3, 41}	43.~~Ra	33, 41, MT
{3, 41}	44.Ra	43, DNE
{3, 41}	45.~Pa & ~Qa	41, 42, &I

{3, 41}	46.(~Pa & ~Qa) & Ra	44, 45, &I
{2, 3, 19, 41}	47.((~Pa & ~Qa) & Ra) & ~((~Pa & ~Qa) & Ra)	23, 46, &I
{2, 3, 19}	48.~~Pa	41, 47, RAA
{1, 2, 3, 19, 20}	49.~Pa & ~~Pa	40, 48, &I
{1, 2, 3, 19}	50.~~Sa	20, 49, RAA
{1, 2, 3, 19}	51.Sa	50, DNE
{1, 2, 3}	53.~Ta ⇾ Sa	19, 51, CP
{1, 2, 3}	54.Sa ⇿ ~Ta	18, 53, ⇿I
{1, 2, 3}	55.∀x[Sx ⇿ ~Tx]	54, UI

The Refutation of Compatibilism

∀x[Sx ⇿ ((Px & Qx) & ~Rx)], ∀x[Tx ⇿ ((~Px & ~Qx) & Rx)], ∀x[Rx ⇿ ∃y[Fxy & Uy]], ∀x[∀y[∀z[(Uz & Gxy) ⇾ (Fyz ⇿ Fxz)]]] : ∀x[∀y[∀z[(Gxz & Gyz) ⇾ ~(Sx & Ty)]]]

{1}	1. ∀x[Sx ⇿ ((Px & Qx) & ~Rx)]	P
{2}	2. ∀x[Tx ⇿ ((~Px & ~Qx) & Rx)]	P
{3}	3. ∀x[Rx ⇿ ∃y[Fxy & Uy]]	P
{4}	4. ∀x[∀y[∀z[(Uz & Gxy) ⇾ (Fyz ⇿ Fxz)]]]	P
{ - }	5. ∀x[Rx v ~Rx]	T[39]
{6}	6. Gac & Gbc	A for CP
{ - }	7. Rc v ~Rc	5, UE
{8}	8. Rc	A for vE
{3}	9. Rc ⇿ ∃y[Fcy & Uy]	3, UE
{3}	10.(Rc ⇾ ∃y[Fcy & Uy]) & (∃y[Fcy & Uy] ⇾ Rc)	9, ⇿E
{3}	11.Rc ⇾ ∃y[Fcy & Uy]	10, &E
{3, 8}	12.∃y[Fcy & Uy]	8, 11, MP
{13}	13.Fcd & Ud	A TD
{14}	14.Sa & Tb	A for RAA
{14}	15.Sa	14, &E
{1}	16.Sa ⇿ ((Pa & Qa) & ~Ra)	1, UE
{1}	17.(Sa ⇾ ((Pa & Qa) & ~Ra)) & (((Pa & Qa) & ~Ra) ⇾ Sa)	16, ⇿E
{1}	18.Sa ⇾ ((Pa & Qa) & ~Ra)	17, &E
{1, 14}	19.(Pa & Qa) & ~Ra	15, 18, MP

[39] Again, this is introduced as a theorem to save time.

{1, 14}	20.~Ra	19, &E
{3}	21.Ra ⇿ ∃y[Fay & Uy]	3, UE
{3}	22.(Ra ⇾ ∃y[Fay & Uy]) & (∃y[Fay & Uy] ⇾ Ra)	21, ⇿E
{3}	23.∃y[Fay & Uy] ⇾ Ra	22, &E
{1, 3, 14}	24.~∃y[Fay & Uy]	20, 23, MT
{25}	25.Fae & Ue	A for RAA
{25}	26.∃y[Fay & Uy]	25, EI
{1, 3, 14, 25}	27.(∃y[Fay & Uy]) & (~∃y[Fay & Uy])	24, 26, &I
{1, 3, 14}	28.~(Fae & Ue)	25, 27, RAA
{1, 3, 14}	29.∀y[~(Fay & Uy)]	28, UI
{1, 3, 14}	30.~(Fad & Ud)	29, UE
{4}	31.∀y[∀z[(Uz & Gay) ⇾ (Fyz ⇿ Faz)]]	4, UE
{4}	32.∀z[(Uz & Gac) ⇾ (Fcz ⇿ Faz)]	31, UE
{4}	33.(Ud & Gac) ⇾ (Fcd ⇿ Fad)	32, UE
{6}	34.Gac	6, &E
{13}	35.Ud	13, &E
{6, 13}	36.Ud & Gac	34, 35, &I
{4, 6, 13}	37.Fcd ⇿ Fad	33, 36, MP
{4, 6, 13}	38.(Fcd ⇾ Fad) & (Fad ⇾ Fcd)	37, ⇿E
{4, 6, 13}	39.Fcd ⇾ Fad	38, &E
{13}	40.Fcd	13, &E
{4, 6, 13}	41.Fad	39, 40, MP
{4, 6, 13}	42.Fad & Ud	35, 41, &I
{1, 3, 4, 6, 13, 14}	43.(Fad & Ud) & ~(Fad & Ud)	30, 42, &I
{1, 3, 4, 6, 13}	44.~(Sa & Tb)	14, 43, RAA
{1, 3, 4, 6, 8}	45.~(Sa & Tb)	12, 13, 44, EE
{46}	46.~Rc	A for vE
{3}	47.∃y[Fcy & Uy] ⇾ Rc	10, &E
{3, 46}	48.~∃y[Fcy & Uy]	46, 47, MT
{49}	49.Sa & Tb	A for RAA
{49}	50.Tb	49, &E
{2}	51.Tb ⇿ ((~Pb & ~Qb) & Rb)	2, UE
{2}	52.(Tb ⇾ ((~Pb & ~Qb) & Rb)) & (((~Pb & ~Qb) & Rb) ⇾ Tb)	51, ⇿E
{2}	53.Tb ⇾ ((~Pb & ~Qb) & Rb)	52, &E

{2, 49}	54.(~Pb & ~Qb) & Rb	50, 53, MP
{2, 49}	55.Rb	54, &E
{3}	56.Rb ⇿ ∃y[Fby & Uy]	3, UE
{3}	57.(Rb ⇾ ∃y[Fby & Uy]) & (∃y[Fby & Uy] ⇾ Rb)	56, ⇿E
{3}	58.Rb ⇾ ∃y[Fby & Uy]	57, &E
{2, 3, 49}	59.∃y[Fby & Uy]	55, 58, MP
{60}	60.Fbe & Ue	A TD
{4}	61.∀y[∀z[(Uz & Gby) ⇾ (Fyz ⇿ Fbz)]]	4, UE
{4}	62.∀z[(Uz & Gbc) ⇾ (Fcz ⇿ Fbz)]	61, UE
{4}	63.(Ue & Gbc) ⇾ (Fce ⇿ Fbe)	62, UE
{60}	64.Ue	60, &E
{6}	65.Gbc	6, &E
{6, 60}	66.Ue & Gbc	64, 65, &I
{4, 6, 60}	67.Fce ⇿ Fbe	63, 66, MP
{4, 6, 60}	68.(Fce ⇾ Fbe) & (Fbe ⇾ Fce)	67, ⇿E
{4, 6, 60}	69.Fbe ⇾ Fce	68, &E
{60}	70.Fbe	60, &E
{4, 6, 60}	71.Fce	69, 70, MP
{4, 6, 60}	72.Fce& Ue	64, 71, &I
{4, 6, 60}	73.∃y[Fcy & Uy]	72, EI
{2, 3, 4, 6, 49}	74.∃y[Fcy & Uy]	59, 60, 73, EE
{2, 3, 4, 6, 46, 49}	75.(∃y[Fcy & Uy]) & (~∃y[Fcy & Uy])	48, 74, &I
{2, 3, 4, 6, 46}	76.~(Sa & Tb)	49, 75, RAA
{1, 2, 3, 4, 6}	77.~(Sa & Tb)	7, 8, 45, 46, 76, vE
{1, 2, 3, 4}	78.(Gac & Gbc) ⇾ ~(Sa & Tb)	6, 77, CP
{1, 2, 3, 4}	79.∀z[(Gaz & Gbz) ⇾ ~(Sa & Tb)]	78, UI
{1, 2, 3, 4}	80.∀y[∀z[(Gaz & Gyz) ⇾ ~(Sa & Ty)]]	79, UI
{1, 2, 3, 4}	81.∀x[∀y[∀z[(Gxz & Gyz) ⇾ ~(Sx & Ty)]]]	80, UI

3. The Spatial Strategy for Idealism

The Significance of Space

Given that spatiality, quantitativity and objectivity are coextensive, we can see that anything that is not spatial must be qualitative and subjective, and thus mental. This may explain why, historically speaking, a popular strategy for idealists has been to attempt to show that fundamental reality cannot be spatial, though the idealists who have pursued this idea may not have thought of it in these terms.

Before I go on to my own arguments against physicalism, which run in a slightly different direction, it will be worth our while to pause and consider two versions of the spatial strategy, as both will introduce us to concepts that will become useful later on. The two versions I shall examine are Kant's and Bradley's.

These are, on the face of it, two very different arguments. Kant argues that the appearance of spatiality must be something created by the mind - it cannot exist in whatever causes our ordinary experiences. Bradley argues that fundamental reality must be a unity, without any real relations within it. However, they agree on a point of basic intuition; that at least some of the relations we perceive between objects are not 'real' in the sense of belonging to the ontological fundament.

Kant and Bradley both have a reputation for dense and confusing philosophical writing (not wholly deserved; Kant may well have been poorly served by his translators, and some of Bradley's writings are excellent), as well as for difficult and counterintuitive ideas. I have done my best in what follows to take them on their own terms, but to do so in the briefest and clearest way possible.

As a final point of clarification, we should highlight something we have not really touched on in the preceding chapter, which is the relation of *time* to space. As we shall see later, contemporary physics tells us the two are so closely linked as to be interchangeable in certain special circumstances. As a result, I think it is appropriate to read 'spatial' henceforth as 'spatiotemporal'; arguments about the reality (or otherwise) of time thus have as much bearing on our question as arguments about space[40].

---

[40] Because a thing by definition cannot be spatiotemporal unless it is both spatial and temporal. Thus, just as a non-spatial thing must be mental, so must a non-temporal thing. Remember that we are concerned only with *genuine* spatiality and temporality; the mental may have the appearance of spatiality, as in the way bodily pains have apparent approximate locations, or of temporality, as in the way experiences seem roughly ordered, provided it is not possible to exactly specify the spatial or temporal dimensions or relations of them. And we can see that the

The Transcendental Aesthetic

I find it quite hard to introduce Kant's arguments impartially. As an empiricist and radical nominalist, I have very little patience for the guiding idea of Kant's 'critical philosophy', the synthetic a priori. There isn't time or space here for me to enter the murky waters of this controversy, though, so I shall do my best to put it aside and allow that there are synthetic a prioris.

Kant's arguments about space appear in the section of the Critique of Pure Reason entitled the 'Transcendental Aesthetic'[41]. Here, his goal is to show that space and time are what he calls 'pure intuitions' – mental phenomena which are representations, but of the *form* (structure) rather than *matter* (content) of experience. This, he claims, will form a basis for synthetic a priori judgements, because pure intuitions can be known a priori (that is, they can be known without knowing anything about the content of an experience), and yet what is known is not analytic (that is, what is known through pure intuition is not contained purely in the concepts deployed, because pure intuitions deploy no concepts; nor is it known because its converse is contradictory).

Kant's argument has three stages: first, he argues that there must be pure intuitions. He then divides arguments about space from those about time and gives a 'Metaphysical' and then a 'Transcendental' exposition of each. A 'metaphysical exposition' in this case is an argument which shows that the phenomenon in question must be a pure intuition because of how we relate to it; a transcendental exposition shows how other things knowable a priori depend on the phenomenon (thereby implying that the phenomenon is knowable a priori since no a priori knowledge can, by definition, depend on anything a posteriori). Barring some problems with the organisation of the text which Kant himself admits at B48, the metaphysical expositions of space and of time are similar enough that they may be considered together; the transcendental expositions bear separate study.

The Existence of Pure Intuitions

Shabel (2010) identifies two arguments in the first stage of the Aesthetic; a hylomorphism argument and a thought experiment. The hylomorphism (separation of

---

mental can never fulfil this requirement because it is also non-quantitative. Spatial and temporal intervals can be specified exactly only in terms of some non-mental standard, but as we have seen it is not possible to clearly and absolutely relate subjective phenomena to such a standard.

[41] All references to the Critique are to the Guyer and Wood translation (1998); I have used the second edition of the Critique, although for our purposes the differences between the two versions are minor.

matter/content from form/structure) argument (which appears at B34) works by claiming that the form of experience must be 'known' (in at least some loose sense of the term) before its content can be apprehended. As Kant's understanding of the a priori/a posteriori divide relates only to the content of experience (i.e. something that is known without input from experiential contents is known a priori), this means the form of experience must be known a priori.

By 'form', or 'structure', of experience, Kant means the spatiotemporal arrangement of contents, which makes this argument a little stronger than it prima facie appears. The idea is that if all our experiential contents were not separated by spatial and (especially) temporal relations[42], they would form an unintelligible mass and we would be unable to relate to them in the useful, significant way that we ordinarily do. Therefore, this spatiotemporal arrangement must, Kant thinks, precede its content into the mind, not temporally (because temporality is a feature only of the arrangement[43]), but in a quasi-ontological way – any judgement made about the content of an experience requires a judgement about its place in the structure.

The second argument in Kant's first stage, which appears at B35, is much weaker. Kant's thought experiment is to ask us to subtract from any given representation – that is, fully conceptualised experience – of a body two things; first, any concept added to it by the understanding, such as substance or divisibility, and second, any sensible content (roughly, qualia – colour, hardness etc.).

Kant claims that what remains is an extension and a form, a strikingly Lockean sentiment which has the classic problem of all 'primary qualities' claims. This is that it is hard to see why we should not draw our awareness of extension and form from experiential *content* (and thus have it fall into the same category as the qualia already removed). If nothing else, the selection of extension and form seems quite arbitrary, though I accept that much of my hostility to the argument may come from my aforementioned empiricist bias.

---

[42] As we saw in the preceding chapters, there are certainly mental items which are not spatially spread out. However, I do not think this makes a difference to Kant's actual arguments.

[43] Kant does have separate arguments in the metaphysical exposition of the concept of time which he claims show that time is nothing more than a feature of the structure of experience, but as these arguments rely on the possibility of pure intuitions, they cannot be used to shore up the hylomorphism argument for same.

The First Metaphysical Exposition - Experiential Contents

Whatever the strength or otherwise of the thought experiment, Kant takes himself to have demonstrated the possibility of pure intuitions and moves to the task of showing that space and time are such things. He gives a metaphysical exposition consisting of four arguments for each concept[44], but the arguments in each case are basically similar. The arguments are designed to each close off one alternative kind of thing that space and time could be other than pure intuitions.

The first argument is to show that time and space are not empirical intuitions – that is to say, that they are not features of the *content* of experience. The only way to 'create' or apprehend space and time based on the content of experiences would be to build our representation of them out of the relations between experiential contents, but these presuppose spatiality and temporality – we cannot be or become aware of these relations without a pre-existing conception of their spatiotemporal nature[45].

Expressed thus, the argument is clumsy and inconclusive. Buroker (2006) unpacks it in more detail as posing a dilemma for Leibnizian accounts of space and time. On this view, space and time are not fundamental, substantial entities (as is held by the rival Newtonian position), but are constitutively sustained by the set of all spatial relations between (physical) objects. Buroker's reading of the argument says this view must either hold that spatiotemporality is part of the nature of these relations – and therefore they presuppose the spatiotemporality they purport to explain – or spatiotemporality appears by radical emergence, a concept few philosophers are willing to endorse without reservation. In this form, the argument looks rather stronger, though the trade-off for this is a reduction in scope.

The Second Metaphysical Exposition - Abstraction

The first exposition has nothing to say, of course, to an anti-Leibnizian who takes space and time as substantial, but the problem it highlights - what we might call the

---

[44] A fifth 'metaphysical' argument in the case of time is given at B47, but at B48 Kant describes it as 'properly transcendental', and I shall follow the convention of dealing with it as such.

[45] It might be better to say 'we cannot be aware of these relations *as spatial or temporal* without a pre-existing conception of spatiotemporality'. It is only the fact that we can identify some relations as spatial, temporal or both that enables us to know that they are the ones from which space and time should be constructed, that they are the ones capable of being constituents of space as a whole. There is a clear echo here of our holistic treatment of spatiality in the previous chapters – the spatiality of any given element of reality is to be defined in terms of its belonging to some spatial whole or field.

'emergence problem' - is an established issue for the Leibnizian schema. The second exposition takes this point and builds on it; Kant asks us as a thought experiment first to attempt to conceive of space and time in abstraction from objects and then of objects in abstraction from space and time.

The reason the latter is impossible, in the Kantian schema, is because differentiation of the manifold of sensation – the sum total of all our experiential inputs - into objects requires the arranging of its elements spatiotemporally. There simply would be no intelligible objects – nothing discernible as 'an object' – without this arrangement of elements of sensation.

Meanwhile, Kant claims, an empty space is quite naturally conceivable – by itself, it is a possible element of experience, though perhaps not a sensory one. For Kant, the second exposition is a more general argument than the first that judgements about space and time cannot be a posteriori. In underlying structure, it goes along these lines: a posteriori judgements are only possible if the manifold of sensation is broken up into independent elements, because a posteriori judgements fundamentally consist in associating such elements with one another. But we know that the manifold of sensation is broken up spatiotemporally so space and time must be established prior to the making of a posteriori judgements.

This is all well and good, but it presupposes that space and time are not substantial elements of reality. That is, it assumes that there is nothing in the reality that our experiences are experiences *of* which separates them. If sensation arrives as an undifferentiated manifold, as Kant suggests, then yes, space and time must be added to it by experience. But it is very much not our usual way of thinking about our experiences and the world they are caused by to think that experiences come to us in this non-spatiotemporal way.

Our normal way of thinking about the world is, I think, essentially Newtonian; we consider space and time to be features of objective reality, substances or at least substantial in their own right. This is one reason why idealism is often so hard to accept, since as we saw in the last chapter, it requires us to completely overturn this common view of space. With substantial space and time – space and time that are completely independent of us as human subjects – spatial and temporal judgements would be every bit as a posteriori as we ordinarily take them to be.

The Third Metaphysical Exposition - Parts of Space and Time

The third exposition – probably the most difficult of the four to get a good grasp of – is directed exactly at this point. Kant seeks to establish that space and time are not substances by showing that there is a part-whole relationship characteristic of substances which does not hold between space and its parts or time and its parts. The argument is ferociously obscure in the text and complicated by the fact that it has a second purpose: to show that space and time cannot be *concepts* (in Kant's technical sense, which later in the Critique he will call categories).

Here is the third metaphysical exposition of the concept of space as it appears at B39:

> *"Space is not a discursive, or, as is said, general concept of relations of things in general, but a pure intuition. For, first, one can only represent a single space, and if one speaks of many spaces, one understands by that only parts of one and the same unique space. And these parts cannot as it were precede the single all-encompassing space as its components (from which its composition would be possible), but rather are only thought* ***in it****. It is essentially single; the manifold in it, thus also the general concept of spaces in general, rests merely on limitations. From this it follows that in respect to it an a priori intuition (which is not empirical) grounds all concepts of it. Thus also all geometrical principles, e.g., that in a triangle two sides together are always greater than the third, are never derived from general concepts of line and triangle, but rather are derived from intuition and indeed derived a priori with apodictic certainty." (1787 p.175 (B39))*

Even dividing this argument into two purposes, it is hard to bring out how it is supposed to work. The key sentence, I believe, is 'And these parts cannot as it were precede the single all-encompassing space as its components ... but rather are only thought **in it**.' The point being made is that space is not a collection of atomic, independently-existing regions, but *one thing*. When we speak of a specific region, we are performing a

conceptual carving-up of something that in reality is undivided. Implicit in the argument is that this is not how *substances* behave[46].

The implicit claim is reasonable enough; the explicit claim is more problematic. Kant appears to take as justification for it the fact that 'one can only represent a single space'; multiple spaces are always understood as spatially related to one another, and thus ultimately part of the same, wider space. This idea has been under fire for some fifty years now, starting with Quinton's 'Spaces and Times' in 1962; to Quinton's explicit critique of Kant, we can add such diverse theories as David Lewis' modal realism and Hugh Everett's many-worlds interpretation of quantum mechanics.

All these views have in common the – I believe correct – idea that it is at least possible that there should be multiple separate spaces which are not spatially related to one another. While we cannot blame Kant for failing to take account of ideas not developed until long after his death, we must acknowledge that the third exposition can only be seen now as a failure; its single premise is false[47].

The second purpose of the argument goes out with the first, but deserves at least a mention. For Kant, the concepts (objects of the understanding) of individual spaces and space in general are created by, for want of a better expression, thinking divisions into the one unified intuition of space[48]. As such, they depend on – are posterior to – the intuition, and this guarantees that the intuition cannot itself be a concept. This is less important in the context of a general case against matter than in Kant's philosophy, since were space and time concepts, this would still constitute a form of idealism, but it is still significant; it is space's and time's status as a priori intuitions (necessary features of all *human* consciousness) which gives them their objectivity. If space and time were concepts, they would be subjective (at least, according to Kantian metaphysics and definitions) because they would be only internal features; as features of experience, they do relate us to the world of things-in-themselves, albeit mediately.

---

[46] This depends a bit on the sense of 'parts' which is used; I assume, however, that the sense of parthood Kant is making use of is similar to the sense of parthood expressed by the predicate G which we used in chapter 2.

[47] It might be argued that Kant is not claiming that we can't *conceive* multiple spaces, but that we can't *imagine* (in the technical sense of 'mentally visualise') them. It is not clear, however, that we should take the limits of what we can mentally visualise as implying anything about what is possible - this would require the actual limits on what we can mentally visualise to coincide with the logical limits of what we could ever mentally visualise, and while Kant sometimes seems to talk as if this is the case, I cannot think of a good argument for this position.

[48] That is to say, the distinctions that result are purely mental artifacts and not features of intersubjective reality in any way. Of course, if idealism turns out to be true, then this definition will have to be quite delicately reformulated, but that is a question for another time.

The Fourth Exposition - Infinity

The final exposition is perhaps the weakest of the four, even on its own terms. Here Kant is again trying to close off the possibility that space and time might be concepts. His argument is that space and time are *given as* infinite – our awareness of space and time unavoidably includes their infinity. Infinity is a specific magnitude, but general concepts cannot *have* specific magnitudes[49] (because generality and specificity are opposed), so space and time cannot be concepts.

The idea that general concepts cannot have specific magnitudes is at least fair enough, but it is not at all clear that space and time are given as infinite. Certainly, our sensations convey to us no infinite magnitudes (though for Kant a pure intuition is a separate thing from sensation). Kant wants us to accept that we not only can but automatically do represent infinity to ourselves in intuition[50]; I am not sure that we can represent (in the Kantian sense) infinities at all. It may indeed be that space and time are *not* infinite; various scientific ideas in the last hundred years have investigated quite seriously the ideas that there might be real boundaries of space or that time might really have an end, or have begun.

Kant's case, then, is about half-successful. He certainly seems to have delivered a compelling critique of the Leibnizian school of thought on space and time[51], but his attempts to dispense with the Newtonian view (and the possibility that space and time are concepts, a position which perhaps we need not take seriously as it is not clear exactly what it means or what considerations would make it appealing) are less successful. This leaves much work for the two transcendental expositions to do.

The Transcendental Exposition of Space

The transcendental expositions are both designed as general demonstrations that the terms 'space' and 'time' refer only to features of human experience. As such, they hit all attempts to assign space or time a reality or mind-independence simultaneously. The

---

[49] I understand the claim that general concepts cannot have specific magnitudes as meaning that if I ask you 'how long is the concept of (for example) length?' there is no correct answer; the question should be taken as ill-formulated.

[50] It's not clear what Kant thought this intuition of infinity involved, whether he meant an actual infinity, or just a sense of the hypothetical 'for any point I reach, it is logically possible for me to go beyond it in any direction'. Either way, I am inclined to disagree with Kant that this *is* a feature of our intuitions – in the Kantian or any other sense – of space and time.

[51] From a historical point of view, this is likely to have been the philosophy Kant was most familiar with, thanks to the work of Christian Wolff.

structure in both cases is the same: Kant gives some proposition he believes is knowable a priori and shows (albeit briefly) how knowledge of space and time is a necessary condition of it. If each of these steps is correct, then it follows that space or time must be knowable a priori, which means (according to Kant, though all the alternative positions seem to me far less appealing) they must be features 'brought to the table' by the structure of human experience.

It is worth noting at this point that within the mental realm of any subject, time is a more general feature than space, in that all experiences are temporally related to each other, but some experiences – the more diffuse deliverances of 'inner sense', perhaps, such as emotions – are not (seem not to be) spatially related to others. As such, to a certain extent the transcendental exposition of space could be dispensed with and the whole case hung upon the transcendental exposition of time. This would not give Kant quite the result he desires, so I will briefly cover the argument about space first.

The transcendental exposition of the concept of space starts from the claim that the propositions of geometry are synthetic a priori truths. Even further, Kant claims that these propositions are *apodictic* – that special Kantian sense of necessity which means that awareness of the proposition includes immediate awareness of its necessity. When Kant says 'the propositions of geometry', he means the propositions of the Euclidean system; while criticisms of Euclid's *Elements* had existed since its composition, it was not until after Kant's death that non-Euclidean geometries gained widespread intellectual currency. As such, Kant's examples of apodictic geometrical propositions are all now out of date; one example he uses is the necessary three-dimensionality of space, where now mathematicians quite happily deal with theoretical spaces with many more, and even infinite numbers of, dimensions.

Nevertheless, it is not clear that these advances in geometry invalidate Kant's position, since with my admittedly limited knowledge of the field of geometry, I cannot see how they would change the modal status of the best current theory - the best current theory will still consist of truths which purport to be as necessary as those Kant took to be necessarily true, they will just not be the same truths. Certainly, propositions that Kant and his contemporaries took as necessary have been revealed to be contingent – it is logically possible, for example, that we could have lived in a 4-dimensional universe – but other

propositions have arisen to replace them, at least in terms of forming the grounds of geometry as an a priori science[52].

Allowing the premise that there are geometrical propositions which are knowable a priori and apodictic without being analytic, does this justify the assertion that space is a pure intuition? This is a tricky question to answer; if space is a pure intuition, then it certainly forms a basis for a priori geometric judgements. As Kant points out, if space is a concept then geometric propositions grounded in it cannot be synthetic, since a conceptual judgement is limited within its own bounds - it cannot tell us about the application of the concept. If space is a substance, presumably it can only ground a posteriori judgements – but as we have seen, Kant has no successful argument against the substantiality of space.

For me, the problem with this transcendental exposition is that there is no clear link between space and geometry of the type Kant posits. This, perhaps, is something that has only become truly apparent since the development of non-Euclidean geometry, but I believe the problem was at least latent in the relationship between Euclidean geometry and space. Whatever its transcendental status, space is from a pragmatic viewpoint part of reality – we live in it, move in it and relate to other objects through it. Kant himself does not deny this, as we shall see. Pure geometry, by contrast, is an *abstract* subject. Applied geometry is just that – a combination of pure geometry and real space. Applied geometry, therefore, is synthetic but actually a posteriori, since we cannot know a priori which particular set of geometric axioms applies to real physical space. On the other hand, it seems to me that pure geometry is analytic and a priori. The transcendental argument falls down over a confusion between pure and applied geometry.

The Transcendental Exposition of Time

This just leaves the transcendental exposition of time. There are two a priori concepts which Kant claims owe their a prioricity to time's status as pure intuition. The first of these is what I'll call 'the axiom of sequentiality', which is the axiom that for any two moments, if they are numerically distinct, then they cannot be simultaneous. More precisely, moments (or instants or whatever the basic units of time should be taken to be) are individuated purely by their place in a or the temporal sequence[53].

---

[52] Once again, I shall bite my nominalist, Humean tongue and refrain from arguing that the general a priori nature of geometry is at least impossible to prove.

[53] This is a point complicated horrendously by the fact that some recent work in advanced theoretical physics has suggested that time might have two dimensions rather than just one (Bars, 2006). It is for this reason that I have referred to this axiom as that of 'sequentiality' rather than

Kant's second 'concept' for this argument is the concept of change (or motion). This is a tricky concept to pin down, as we shall see in a moment, but a common-sense notion will serve right now. I think the argument that, if sequentiality and change are both synthetic a priori concepts, they must be grounded in the relevant way in time, and therefore time must itself be a priori, is fundamentally valid. Certainly, it is only through an awareness of time that, as Kant puts it, "a combination of contradictorily opposed predicates ... in one and the same object" (1787, p.180 (B48)) is possible.

Unfortunately, there are again problems with the premises. Firstly, I think the axiom of sequentially is not synthetic, but analytic. It seems to me obvious that the concept of an individual moment (point in time) involves the exclusion of its being any other moment, and this is exactly what the axiom of sequentiality means. The fact that moments are individuated by their positions in the temporal sequence or field is part of a correct understanding of the concept of a moment; it is not synthetic, and thus cannot be contributed by a Kantian intuition at all.

The concept of change is a little more complex to criticise. If change as a concept is the concept of an object's having two incompatible properties, it seems to me to be clearly grounded in the content of experience – the only times we make judgements of this type are when experience (whatever mechanism gives it its temporal order) gives us, for example, the judgement 'The light is blue', followed by the judgement 'The light is red'. It is not clear to me what, if any, feature of this sequence can be abstracted and shown to be a priori unless we already knew that time itself was a priori.[54]

It might be thought that change-judgements about the hypothetical or the fictional present a problem for this case. My judgement, for example, that Holmes' pipe changed from empty to full is not grounded in any actual experience I have of Holmes filling his pipe. There are two strategies that we could pursue against this point. The first is simply to point out that that my judgement about Holmes' pipe is still grounded in experience; specifically, my experience of reading Conan Doyle's assertion that Holmes filled his pipe.

This at least solves the issue for fictional judgements – they turn out to be grounded in experience and are thus a posteriori. It is not clear how one would pursue this strategy for judgements about the hypothetical, unless one could defend the substantial and controversial claim that all hypothetical objects are in some way dependent on

---

'linearity', though even 'sequentiality' is far from an ideal term. The point is that, however many dimensions time has, no two moments can occupy exactly the same point in the time-field.

[54] I accept that this sentiment is Humean at least in form, so perhaps my bias has got the better of me at the last, but the question remains, whether Humean or not.

experience[55]. The other approach offers a more general answer. This is to say that our ability to apply the concept of change to the purely hypothetical depends on our having learnt first to apply it to the actual. This, I think, is an entirely plausible claim and deals neatly with the problem.

The role of the concept of change is to "make comprehensible the possibility of an alteration" (1787, p.180 (B48)) – that is, make it possible to think of two different object-states as belonging to the same object. It is, or at least is very tightly bound-up with, the concept of an object's identity-over-time, so clearly it depends on an appropriate understanding of time as a feature of the universe. The point I am making is that nothing a priori can ever cause one to *use* (in the most general sense of the word) such a concept; the only things which can prompt a change-judgement are experiential *contents*[56], so even if there is in some sense[57] an a priori concept of change, it will only ever figure in one's consciousness in an a posteriori fashion.

It seems to me, then, that of the two supposed synthetic a prioris Kant claims as essentially derived from the a priority of time, one is not synthetic and the other is not a priori. The question now is whether the two in combination can show that time is an a priori intuition, and I do not think they can. The concept of change, whatever it is, cannot justify any attribution of a priority; the axiom of sequentiality quite definitely does not show that time is an intuition. Indeed, given how tightly bound up together time and sequentiality are, I am tempted to say that sequentiality suggests all the more strongly that time in fact is a concept[58]; whichever way one goes, though, the possibility is left open that time is real – a substantial feature of reality.

---

[55] Perhaps by arguing that we can imagine no possibility which is not a combination of experiences we have actually had. This has been a popular position at times, but it has also been challenged enough to make it risky now; I do not think we need the claim, so I will not investigate it in any detail here.

[56] Kant takes the view – which in his own time was the norm, though general relativity has now refuted it – that space and time are uniform throughout themselves. Thus, there is no alteration in space or time; nor can there be in concepts, because concepts are general, and particularity is a condition for mutability.

[57] And it could not be a Kantian sense – Kant's technical sense of 'concept' (to which he does not always stick) is equivalent in thought to 'intuition' in experience; i.e. a kind of judgement.

[58] That is to say, the relationship between sequentiality and time is of the concept of one to the concept of the other, thus allowing room for the common-sense idea that the concept of time is derived from the reality of time (its existence as part of the 'real' world).

Empirical Reality

Before we move on to Bradley and absolute idealism, it's worth pausing to consider a point implicit in the transcendental aesthetic which may be Kant's most significant contribution to the history of idealism. It's an idea he develops in greater detail elsewhere, but his arguments for it fundamentally come down to the transcendental aesthetic, and it's an idea that will be fundamental to our progress beyond the spatial strategy.

The idea in question is Kant's famous distinction between phenomena - things as they appear to us - and noumena - things as they are in themselves. More importantly, Kant identifies that under certain conditions, as for example if space and time are pure intuitions and features of human experience rather than of reality separate from human experience, the phenomenon-noumenon distinction serves to characterise distinct and separate worlds, one consisting of phenomena, arranged spatiotemporally, and one consisting of noumena[59].

In Kant's scenario (though there are other ways of demonstrating such a separation, some of which we will pursue later), the phenomenal (or empirical/experiential) world is reserved for humans and any other species which possesses a human-like set of pure intuitions. The noumenal world has no such limitation; for example, Kant argues that God cannot possess human pure intuitions (because He is omnipresent and atemporal), and so must inhabit noumenal reality[60].

The framework of a phenomenal/noumenal world distinction (I will, by preference, use the term 'experiential' for the phenomenal world, and variously 'fundamental' or 'transcendental' for the noumenal) will be key to our work going forward, because much will turn on whether we can characterise the transcendental world or investigate how it relates to the experiential. Also, we can be interested in the transcendental world for quite different reasons to our interest in the experiential, and determining which world various philosophers and theories are about will clear up a lot of confusion in the remainder of our debate.

---

[59] If, indeed, noumena are non-spatial as Kant assumes. This is a question to which we'll return at the end of this chapter.

[60] Kant actually brings up this point (1787 p.191 (B71-72)) as a further argument that space and time cannot be conditions of things in themselves, because otherwise they would limit God in the same way that they limit us, but while we should no longer find this argument compelling, we can still allow the point about God's place in reality.

Absolute Idealism and the Relational World

We can now move on to an examination of the absolute idealism of F.H. Bradley. A thorough historical survey of idealism should at this point touch on Hegel, as the bridge between Kant and the British idealists of whom Bradley is often held up as archetype, but Hegel's work is voluminous and legendarily difficult, so for brevity's sake I have omitted him, along with the continental tradition he inspired. Instead, I will focus on Bradley, as well as one of his key forebears, T.H. Green.

It is worth mentioning here that the analytic philosophy of the 20th century is generally thought to have been founded upon the corpse of this form of idealism, it having been refuted conclusively by Russell and Moore sometime between 1900 and 1925. While it is certainly true that Russell and Moore did at the last argue Bradley into the ground, and that their critique of absolute idealism made many significant and important points, awareness is starting to emerge that this 'refutation' may not have been as complete, conclusive, or academically rigorous as generally thought[61]. As such, the arguments of the absolute idealists bear re-examining.

While the bulk of the argument in this chapter will be drawn from Bradley's work, the overall schema of his philosophy is complex and idiosyncratic even by the standards of 19th-century idealism, though he is often taken as its greatest representative. As such, I feel it will be of benefit to start by looking at the schema, if not the specific arguments, of one of the movement's founders, T.H. Green. In this, I am following Hylton's (1990) treatment of the subject.

Green inherited from Hegel two distinct doctrines which can be seen as something of a legacy from Kant. One of these, discussed in the previous section, has to do with the possibility of differing levels of reality. By the height of British idealism, this view had matured into the coherence theory of truth – the view that a statement is true to the degree to which it agrees with other statements (or the number of other statements with which it agrees). On this theory, the fundamental truths agree to a certain degree with every statement with any degree of truth; different levels of reality are in effect made up of truths which share a degree of coherence[62].

---

[61] For an excellent and thorough discussion of this issue, both historically and philosophically, see Candlish (2007).

[62] That is to say, if a reality is made up of 'statements' – at this stage it would be presumptuous to disambiguate between 'facts', 'propositions' and 'sentences' – then a given level of reality will be made up of a set of statements of approximately the same level of truth; that is, a set

The second key legacy of Kant and Hegel is that the prototypical British idealism is a critique of judgement; it holds that the immanent world – the level of reality public to human beings – is constructed by human judgement-processes (rather than, as for example in Berkeley, by a transcendent or divine process). This is not to say that the British idealists accept wholeheartedly the Kantian schema of two distinct realms, the phenomenal and the noumenal, and is it unclear whether there was any general agreement among them about what precisely, if anything, lay behind the immanent world; they did, however, generally agree that the main world-creative process was human and psychological.

The Relational Argument

Something which became explicit with the British idealists and which was perhaps only implicit in Kant was what could be called 'the relational argument'. This argument, which may be Green's major contribution to metaphysics, starts from the premise that immanent reality (the reality we inhabit most immediately[63]) is inherently relational – it consists of distinct objects standing in relations to one another. This seems like a reasonable enough claim; I know of and can foresee no significant challenge to it.

The relational argument then claims that relations cannot be part of ultimate reality – that, in fact, ultimate reality must be some sort of unified[64] whole. Hence, this form of idealism is sometimes called absolute monism. Obviously, this is a much more controversial premise than the first, but grant it, and the conclusion that the immanent world is ultimately unreal follows by a simple modus ponens. The problem with Green's philosophy, and the reason I do not intend to spend a great deal more time on it, is that Green gives very little argument in support of this second premise.

Bradley, on the other hand, gives a brilliant and much-misunderstood argument against the reality of relations. Before we can go through this in detail, a couple of remarks remain to be made. The first is to raise a key difference between Green's and Bradley's schemas, one which sets Bradley apart from most of his idealist colleagues. Green and Bradley are both in a sense absolute monists, in believing that the immanent world is an artificially-pluralised version of a single unified whole, but Green seems to believe, at least

---

of statements consistent with one another and with a constant number of statements from outside the set.

[63] This is a complex concept to pin down, and we shall devote much more time to it in chapter 5.

[64] Here and for the rest of our discussion of the British idealists, it is important to take 'unified' in the sense they meant the term, which is as meaning 'lacking internal divisions or relations'.

implicitly, that human minds stand outside this unity as a separate category, while Bradley believes that minds as well as bodies ultimately disappear into the absolute. This can make Bradley's system feel rather counterintuitive, and may lead to an impossibility[65], but I do not think that it affects his arguments against relations; only the system he proposes to explain the absolute.

The final point to make before we embark on a study of Bradley's arguments is that Bradley is concerned to establish only that relations are not part of *fundamental* reality. He does not deny that the world we experience is relational – far from it – and he does not deny the utility of the relational way of thinking and talking about the world. He simply aims to show that there is no coherent way to treat relations as *ultimately* real. He takes it as obvious that anything which is ultimately real cannot depend on anything else for its existence – real things must be ontologically independent. This is a vital point, since his argument against relations involves showing that they cannot be ontologically independent[66]. I am personally sympathetic to the idea, though I cannot think of a terribly convincing argument either way on the point.

In the course of his philosophical career, Bradley gave many arguments against the reality of relations, particularly in his later years when he championed idealism against Russell's and Moore's attacks. The running-battle nature of this latter phase makes much of the writing Bradley produced difficult and somewhat disorganised, particularly since the debate involved a good deal of talking at cross purposes and perhaps even outright misrepresentation. As such, I intend to focus mainly on Bradley's masterpiece, *Appearance and Reality* (1908), originally published before this debate began. The main argument against the reality of relations in this book appears in the startlingly-brief third chapter, and it is this argument I will set out here.

Bradley's argument at this point is not so much specifically against the reality of relations, but against the possibility of absolute pluralism. Bradley wants to show that ultimate reality is a unity and therefore involves no relations – and also no properties to be related. The categories of property and relation belong only to appearance – the immanent

---

[65] The impossibility being that minds are supposed to generate the relations which break up the absolute, but minds only exist as things already broken from the absolute – so what broke them? There is not time or space to consider this paradox and its significance, if any, for Bradley's philosophy, but it bears mentioning.

[66] The point is also important because it makes absolute monism almost inevitable; if this independence criterion is to be accepted, any but the narrowest conceptions of independence will lead us to admit that only everything, taken together as a unity, can be independent in the desired sense.

world – and not reality. As such, Bradley essentially gives four arguments; the first showing that properties cannot exist without relations, the second that properties are incompatible with relations, the third that relations cannot exist without properties to relate and the last that relations are incompatible with properties. It is only in the last of these that Bradley deploys his famous regress.

It should be noted here that Bradley clearly has in mind property-instances rather than 'general properties' (what we'd now call universals). I assume that this means Bradley was ultimately some form of nominalist (the idea that predicates are human constructions certainly fits better with his ontological views than any form of Platonism), and held that general properties are derivative of their instances rather than vice versa, but this is not significant for our discussion. What is significant is the need to remember that 'properties' here means property-instances.

The arguments fall into two obvious pairs, the first relating to properties and the second to relations. It might be thought that there is some redundancy between the second members of each pair, since both essentially aim to show the same thing, but the arguments Bradley uses to get to that point in each case are clearly and distinctly different, and both bear close scrutiny. I'll take the arguments in the order they appear in the original text.

Properties Without Relations

Bradley's first argument, that properties cannot exist[67] without relations, is based on the claim that there are certain relations – primarily those of identity and difference - which are necessary to the formation of any kind of property-judgement. In this, we can clearly see an echo of Kant's pure intuitions and his idea that the content of experience, which arrives from the noumenal realm, comes to us as a unity. Bradley's basic point is that any time we formulate the judgement 'such-and-such is a property', judgements about certain relations of such-and-such are implicit.

It is important to remember the Kantian (and absolute monist) background to this argument. According to this background, divisions and plurality are only added to sensation

[67] Bradley assumes that everything that is possible is conceivable, and his argument might be more precisely described as showing that properties are *inconceivable* without relations; Bradley would regard the two formulations as interchangeable. Not everyone would agree that the limits of the possible and the conceivable coincide – for my own part, I can see no clear reason for rejecting the claim that there are possible objects which would elude my conception even by direct acquaintance – but I believe the points Bradley makes, properly understood, penetrate down from the conceptual to the ontological level.

when it is processed after reception by a perceiver. Seen in this light, the argument starts to look more comprehensible. When one picks out a property, for example by saying 'this is a patch of redness', one picks it out *as in some way separate from* its neighbours.

It is not the redness that is problematic here, but the 'this'. The point is not that to identify something as a patch of redness requires it to have no red neighbours; we can, after all, imagine indicating some small area of a red sheet and saying of it 'this is a patch of redness'. We must, however, identify it as *this* patch rather than any of its neighbours. Implicit in 'this is a patch of redness' is a distinction between the indicated patch and all others. That relation of distinction is required for the property-judgement to be meaningful – in fact, for it to be a judgement about anything at all.

It is tempting for the realist to object to Bradley here on absolute monist grounds, by saying that surely in the absolute unity he proposes, there are properties not standing in relation to one another, but unified. This is a misunderstanding, says Bradley, because while it is true that there are no relations in the absolute unity, there is also nothing that can properly be called a property, for precisely the reason just given; identifying some phenomenon as a property entails a relation between it and other things.

Bradley goes on to consider whether an 'abstracted property' could count as a property existing without relations. His understanding of abstraction is, I think, that it involves stripping off from a conception of a thing everything related to it (and every relation it stands in)[68]. As he points out, this implies that the term being abstracted – in this case, a property – must have started in a relational situation.

To show by abstraction that properties are real, then, the realist would need to show how the abstracted property *could* be apprehended without starting in a relational situation, even if by a process of apprehension of which human beings are incapable. The process of abstraction by which we attain our unrelated property can only work on something which started out relational; therefore, the unrelated property achieved in this way cannot be independent of relations, even if there is some unrelated property out there which is. Bradley rounds this stage of the argument off with the claim that no other means

---

[68] So, for example, 'redness' would be reduced to the raw feel of redness – something which is ex hypothesi impossible to define, since to define it would be to relate it to something (a general concept, possibly). Remember, Bradley's problem is not about general properties, but particular instances of those properties. The point is that the term 'this redness' can only meaningfully pick out something in the world if that something is distinguishable as *a thing not numerically identical to its neighbours* (or at least some of them).

than abstraction exists for producing unrelated properties – this claim is open to challenge, but I can't think of one.

Finally on the first point, Bradley highlights the perhaps-facile point that plurality of properties – which is after all a precondition of the kind of abstraction just discussed – implies relation, because it implies diversity, which *just is* the standing of objects in relations of difference to one another. If this point is facile, it is only because it fails to engage with the universe that consists of a single property – and even then, such a universe would be one *within which* no ascription of property would be possible (because there's no one within it to make such an ascription – any such claim would have to be made by an external observer).

The fundamental point of the first argument is this: for a thing to be individuated, it must be separate from some other individuals, *unless* the universe is a homogeneous unity, in which case the whole thing can (arguably) be an individual. If there is a plurality, then its members must stand in relations of separation from one another. Thus, a plurality of properties requires at least some relations.

The Compatibility of Properties and Relations

Let us move on to the second part of Bradley's argument; the argument that properties cannot meaningfully coexist with relations – that is, ultimate reality cannot consist of properties in relation any more than it can of properties out of all relation. This argument is a regress, though it is not Bradley's most famous regress argument.

Bradley starts by pointing out that if properties and relations are both real, they must be inseparable, because neither can exist alone - we have just seen that properties cannot exist without relations; Bradley claims, I think reasonably enough, that a relation cannot exist without relating things[69], and assumes that relations can only relate properties. It's open to debate whether substances or abstracta could provide an alternative fundamental relatum to properties (and, indeed, we may be able to dream up other categories to suit), so it may actually be the case that relations can exist without properties, but for now I'll grant Bradley's first step.

If, then, neither relations nor properties can exist without the other, they exist in a state of circular dependence. Properties support relations, because a relation can only exist

[69] Though we might allow that an individual relation can relate other relations, a system composed only of relations could not exist, because at no point would there be a non-relational relatum to 'start things off' ontologically. We shall discuss some scenarios of this type in chapter 4.

as relating something, but properties cannot exist without the support of relations, since it is only by standing in relations of difference to neighbours that any property may be picked out as such. A property therefore performs two ontological roles, as support for and as supported by the relations in which it stands; I shall call these the categorical and relational roles respectively. These two roles constitute separate aspects of the property, which can only both be said to be aspects of the property if they stand in relations either to each other, or to it, or both.

Thus, a property is not a unity, in Bradley's strict non-relational sense, but a composite of two distinct aspects - it is 'internally diverse'. And since its internal diversity implies the existence of a relation between its twin aspects, prima facie at least, these aspects must themselves then bifurcate, into categorical and relational aspects[70]. This implies further relations and so on ad infinitum; if properties and relations are indeed mutually dependent, then properties are revealed to have a fractal, infinitely dividing internal structure. Nor are there any grounds to claim that the regress involved is not vicious; the goal is to explain how the situation of ontological interdependence between properties and relations exists, but the argument shows that any attempt to explain it will end up being relational – that is, invoking at least half (and, granting the interdependence, both halves) of what we hoped to explain.

A first criticism of this argument might come in at the first point, where we have already seen that Bradley does not cover all his bases. This point is the claim that relations can only relate properties. The Humean rejection of Lockean substance (bare, property-free substrate) is implicit in Bradley's approach to ontology; there is no consideration given to the idea that anything besides properties may stand in relations. But consider this; if a relation could exist between substances (which did not require the relation for their ontological support), it could exist without properties.

Obviously, the antecedent of this conditional is a rejection of a fundamental thesis of absolute monism, but that thesis – monism – is precisely what Bradley is arguing for at this point, so he is not entitled to assume its truth. However much the Lockean theory of substance has been criticised, then (and I count myself no friend to the doctrine), we must examine whether it can be unpacked in such a way as to undo Bradley's argument here.

---

[70] Arguably, this requires that 'aspects' be understood as meaning 'substantial parts' - Bradley is implicitly treating them in a way analogous to parts at the very least. Do two logically distinct aspects of an object necessarily have to be related (in the Bradleyan sense) through the object? Perhaps not, but this would be too big an issue to get into here.

If a relation can relate substances (and the only requirement for this to be the case would be the existence of substances, plural), then it is possible for a relation to exist that does not relate properties. It doesn't become possible for a relation to exist without relata, though; a relation must still relate *something*, even if what it relates are bare substantival particulars rather than properties. Thus, the dependence of a particular individual relation on its terms is not broken.

This means that when a relation does relate properties – and even if it relates a substance *to* a property – the same dependence problem is going to occur. A relation would not be interdependent with substances, because substances can (at least conceivably) exist by themselves. As soon as a property enters the equation, however, the relation has at least one foot on a cloud, and cannot stand. The property cannot ontologically satisfy the relation the way a substance would.

Thus we see that a substantivalist approach cannot save properties from this regress[71]. There is a more serious problem with Bradley's argument, though, when we remember that it is addressed specifically to *properties*, and not to both properties and relations at once. This is that it is not entirely clear why we should accept Bradley's assertion that the categorical aspect of a property bifurcates in being related.

An opponent of Bradley may wish to argue that, by definition, the categorical aspect of a property is equipped to support relations, though the support of any specific relation is not essential to it. This would prevent the regress from affecting properties at all (though it is not clear where it would leave relations, as we shall discuss in the next section). Whether this position can be unpacked without straying into the problem of properties without relation discussed in the previous section is hard to say, but the option certainly remains open.

Bradley's Critique of Relations

Bradley turns next to the question of whether relations could ever exist without properties – that is, given the problems with substantivalism discussed in the previous section, whether relations themselves could be substances. The answer is a conclusive

[71] The best we could manage would be a class-nominalist account of properties, according to which a general property is the set of substances that instantiate it. This concedes at least a technical victory to Bradley, since it makes general properties ontologically derivative, which is part of what he's aiming at. The status of property-instances becomes rather unclear under class-nominalism; presumably they are that in virtue of which substances possess certain properties, but there's a distinct risk of circularity here, and class nominalism is no less controversial a hypothesis than the substantivalism it is here shoring up.

negative. His arguments on this point cover both the possibility that relations could exist without properties and the possibility that both relations and properties could exist.

We have already seen Bradley rejecting out-of-hand the possibility that a relation could exist without relata, and I think this is fair enough. It is in the nature of a relation to relate; nothing that does not relate can be a relation, and indeed nothing for which it is merely possible that it might not relate can be a relation. While this does not require that a relation always relate the same two terms, it is hard to see how one or other term could be replaced, yet the relation remain numerically the same[72]. Therefore, I think, we may take it, in accord with Bradley, that relations *essentially* relate their terms. This is not, of course, to say that all terms have their relations essentially (though this view is sometimes attributed to Bradley).

However, if relations essentially relate their terms[73], they are in a sense ontologically dependent upon them – they cannot exist without their terms. Remember that Bradley is in a sense a foundationalist; he is seeking an ontology consisting only of things which are independent of any other thing. Thus, relations which essentially relate their terms cannot be considered fundamentally real – and neither can anything (for example, properties) which is ontologically dependent upon them.

Bradley's argument against relations can be seen as a conditional disjunction; 'if relations are real, then either they essentially relate their terms or they do not; but if they essentially relate their terms, they are not ontologically independent enough to qualify as real; and if they do not essentially relate their terms then they are not properly relational'. It is only to shore up the second half of this disjunction that Bradley offers his most famous regress argument, which I will now briefly summarise.

Bradley says that what is most correctly called a relation is a kind of fact – what we might call a relational situation – consisting of two (or more) terms *as they are in relation*,

---

[72] By which I mean, if we imagine a relation R relating terms a and b, we would not normally take it that b could be replaced by c; we would take it that aRb and aRc are different relations – that actually we are talking about aR^1b and aR^2c; different instances of the same *type* of relation, to be sure, but ultimately different entities. We individuate relations on the basis of their terms; different terms mean a different relation. If a man divorces and remarries, we do not talk of the same marriage to a second wife, but of a second marriage; it is hard to imagine a change of wife occurring without it being 'a second marriage' regardless of divorce (or widowing).

[73] Care is needed here; as with properties, I think Bradley is best understood as using 'relations' to refer to relation-instances rather than general relations. Even then, it is at least conceivable that a given relation-instance *could have* related different terms (though Bradley is sometimes accused of holding the contrary) - from our previous example, a man's first marriage *could have been* to a different woman. The point is moot; all that matters is that relation-instances cannot be had without some terms which they relate.

and the relationship in which they stand. If the relation does not essentially relate its terms, then something must explain *why* this particular relationship happens to attach to these particular terms – and this can only be explained by invoking a further relation. This is where the regress sets in.

There has been some controversy, dating back to Russell's critique of Bradley, over whether this regress is indeed as vicious as Bradley thinks, but I believe it clearly is. One sets out to explain relationality; acknowledging that an essentially-relating relation lacks the ontological independence and therefore reality to do the job, one proposes relations which do not essentially relate their terms; but to then explain how these relations actually relate, one finds oneself having to invoke more relations, each of which faces the same problem.

All this applies to relational complexes as wholes. But within a relational complex, even Bradley would have to admit, different parts play different roles. Each property and relation adds something to the whole which is not brought to the table by the others. It might be argued, then, that these elements, which we might call 'additional content', can constitute the reality of relations and their terms.

This train of thought is unproductive, however: first, in respect of relations, it runs straight into the problem already discussed about being essentially relational - if the 'real' part of the relation is essentially relational, then it cannot be ontologically independent, and if it is not then it's not clear that we're talking about a relation or relational phenomenon at all. The problem in the case of this approach to properties is a little more subtle.

If we take the non-relational component of a property as its real, fundamental part, there is nothing to distinguish it from its neighbours - as Bradley points out very early on, anything which is a distinct object in its own right must be distinguished from its neighbours by some form of relation. A fundamental reality composed solely of such relation-free properties would not be diverse and plural at all. In effect, this last attempt to avoid Bradley's arguments ends up granting his point.

The Problem of Foundationalism

The last weakness of Bradley's case, then, is his foundationalism. The whole argument relies on the postulate that nothing which has any kind of ontological dependence on another thing is fundamentally real. This is a hard point to argue, either for or against. It is certainly true that in our everyday life, we experience as real many things

which are ontologically second-order; it might at the furthest extreme be argued that anything composed of atoms is ontologically dependent on those atoms, or that a table is ontologically dependent on its top and legs. We do not ordinarily think tables are unreal, though of course as an idealist this is in a sense precisely what Bradley is asking us to do.

The question really comes down to one's understanding of the term 'fundamental' and its use. Bradley clearly thought that 'fundamental' implied that an object relied on nothing of a lower ontological order for its support – because the fundamental is the lowest ontological order. This is fair enough, but a question remains over whether something can depend on another object of the same ontological order and yet be fundamental. Bradley's work seems to put properties and relations at the same ontological level – it is hard to see how it could be otherwise given that he claims they are mutually dependent – so it must be possible for an item to depend on another of the same level. Can this kind of dependence then occur at the fundamental level?

At the very least, there is an epistemological issue to be considered here – how would we tell if it were impossible that such dependence occur at the fundamental level? No empirical test could get past the world of appearance, and it is difficult to see what argument might bear on the matter, unless we incorporate into our definition of 'fundamental' a clause about independence. To do so may be linguistically justified, but then a question arises as to whether it is really the *fundamental* ontology of the world we should be interested in, or whether such a fundamentality criterion is too stringent to produce a useful result at the end of the enquiry. What we are interested in, ultimately, in the debate between realism and idealism, is determining what the lowest ontological order that actually exists is; which ontological order is such that there is no lower? By this definition of fundamentality, it is possible that even given Bradley's arguments, relations and properties exist, as mutually interdependent, at the fundamental level.

So we must ultimately see Bradley's argument as a failure, though a compelling and challenging one (it's worth noting that the criticism by which it fails is not among those normally cited as part of the 'refutation' of Bradley's system). Bradley fails to justify his monism, but only at the very last hurdle.

This may be for the best, though, for our purposes. If it were true that fundamental reality consisted of a single absolute unity, it is hard to see how this could be reconciled with the definition of mentality given in chapter 1. Since it is an absolute unity, it would be bereft of internal relations and so have no exactly-specifiable distance relations (thus

making it non-spatial), but it also could not stand in any privileged relation to anything[74]. As such, absolute monism makes a nonsense of our definitions of 'mental' and 'physical'.

The Limits of the Spatial Strategy

Kant's and Bradley's arguments are, whatever their flaws, impressive and sophisticated philosophical landmarks. Along with John Foster's functional arrangement argument, which we will consider in the next two chapters, they represent perhaps the best the spatial strategy - the argument that fundamental reality must be non-spatial, and thus mental - has to offer. However, there is a deeper problem with the spatial strategy than the specific issues which thwarted Kant and Bradley.

I have separated Foster's argument because, being narrower in scope and ambition, it does not fall foul of the limit I am about to describe. This is because Foster's argument concerns only the spatiality of the experiential world, and not (as with Bradley and Kant) the transcendental world.

When applied to the transcendental world, we run into an obvious problem with the spatial strategy. We base our - I think undeniably correct - ascription of apparent spatiality to the experiential world on experiential data; it is only through some sort of appearance of spatiality in our experiential fields that we are able to say the experiential world is spatial. Some of the problems Kant's arguments ran into came from attempting to deny this, though the present issue runs much deeper.

The problem is this: to prove that ultimate reality is non-spatial (and thus mental), we need to do more than show that the spatiality of the experiential world cannot be fundamental. We must be able to show that the transcendental world is necessarily non-spatial, which is a much harder task. It is entirely possible (as we shall see in chapter 5 when discussing Foster's arguments) that transcendental reality could include a spatial field every bit as genuine as the experiential world's, but numerically distinct. An argument that shows, as Kant's purports to, that the spatiality of our experience is unrelated to anything standing outside or behind our experience, has nothing to say to this possibility[75].

---

[74] This actually brings up a point we passed over earlier to do with identity. Any object stands in a privileged identity-relationship with itself, but we do not normally consider every object to be subjective. It would be lazy sophistry to hold this up as an argument for idealism; instead, it is preferable to add a stipulation to our definition of subjectivity that we should not count the identity relation when determining if a given item stands in any privileged relations.

[75] As mentioned previously, Kant does offer the argument that if transcendental reality were spatiotemporal, then God would have to be so too, and God must be unlimited, but this is not a claim we can be expected to take seriously in the present era. Even if God exists, it is easy enough to

Bradley's position seems stronger here, in that he offers a complete refutation of the possibility of relations existing outside of the sphere of human mentality. However, while his arguments might have proved that relations and properties as-they-feature-in-our-experience cannot be features of fundamental reality (and I stand by my conclusion that they have fallen short of this goal), there is nothing in them to guarantee that some sort of diversity of which we have no experience or conception could serve to break up the absolute.

And in this we can see the essence of the problem. I do not believe it is controversial anymore to say that for an argument to prove something about a particular world, it must take some facts about that world among its premises. Even the most die-hard logical realists would, I think, acknowledge that the theorems of their logic systems are fundamentally statements about logic and not about any *particular* world.

If we are to make some argument about the spatiality of the transcendental world, then, we must know some things about the transcendental world[76]. But we have very little in the way of epistemic resources for dealing with it. We cannot experience it, by definition. Kant thought that the noumenal realm was unknowable; Bradley would have considered the absolute unity to be inconceivable to human reason (or at least, impossible to fully and clearly conceive).

We certainly cannot hope to know everything about it, even counterfactually. And thus, for any given argument we might wish to make about the properties of the transcendental realm, we cannot be sure that there is not some fact, unknowable to us, which refutes that argument. Those of us prone to more radical empiricism may even go as far as to suggest that we cannot be sure the familiar logic of experiential reality applies to what stands behind it.

And if we cannot be sure of our logic, how could we possibly hope to be sure of any conclusion we draw from it? The spatial strategy requires us to overcome all these difficulties; as such, I am of the opinion that we cannot expect it to bear fruit. We simply can't know enough about transcendental reality to be sure of any conclusion about it.

---

consider him as existing in space and time, but immortally and omnipresently, rather than atemporally and non-spatially.

[76] We cannot hope to argue that no world is spatial, since in this we would already be contradicted by our acknowledgement that the experiential world is spatial. Our best bet might be to argue that spatiality cannot be ontologically fundamental (for example, by taking a hard-line Leibnizian stance), but this would be a complex and challenging task.

4. Intrinsic Content, Humility and Causal Structuralism

We will move beyond the spatial strategy, and reformulate our approach to idealism in order to circumvent the problems just revealed, in the next chapter. There, we will consider the anti-realist arguments of John Foster. Foster's arguments start from a point (rather different from the starting points of Kant and Bradley) about what counts as a complete description of the (a) world, and to do justice to the issue we must site it correctly in both its historical and modern contexts. To do so, we must range quite widely across debates, and as such this will be quite a big diversion, though unfortunately unavoidable.

Foster's starting point is this: the questions 'What is this thing like?' and 'What does this thing do?' are, prima facie, different questions. They appear to have different answers; the latter is answered (more or less generally) in terms of causal properties, and the former in terms of what Foster calls *intrinsic content properties*[77]. Foster's argument is aimed at two key claims: first, that intrinsic content properties are a separate class from causal properties, and second, that there is no *essential* link between any intrinsic content and any causal property.

The main thrust of his discussion is epistemic, at least on the surface. First, he works from the assumption that the physical realist is committed to the claim that mind-independent physical objects must have intrinsic contents, and shows that if they do, we can know nothing about them. Then he shows that the physical realist *must* be committed to this claim, since it is impossible to account for all phenomena in the real world in terms of purely causal properties. Finally, in the argument we shall consider in chapter 5, he shows that the logical separation between causal and intrinsic properties (and the consequent limits on our knowledge) leads to a form of idealism.

It should be stressed at this point that what is at issue is not just *any* knowledge, but specifically *transparent* knowledge, and this is an obscure and difficult notion to get a handle on, particularly in the way Foster uses it. He offers this example:

> *"Suppose I have a sealed envelope and I know that inside it there is a piece of paper on which someone has drawn a geometrical figure, but I do not know what type. If someone who does know tells me,*

[77] In *The Case for Idealism*, Foster calls his argument the 'topic-neutrality thesis' (part 2, esp. ch.4); in *A World for Us*, it is 'the inscrutability of intrinsic content' (ch.2). The discussions are fundamentally similar, with the later simply being more detailed. In what follows, I shall treat them as one argument.

> *correctly, that the figure is a triangle, his specification of the type is* transparent*. If he tells me, again correctly, that it is an instance of that type of figure whose geometrical properties are discussed in the fourth chapter of the only leather-bound book in Smith's library, his specification is* opaque*. In both cases, the information he provides is, in a sense,* about *the intrinsic nature of the figure. But there is also a clear sense in which the first specification ...* reveals *this intrinsic nature and the second ... does not. Unless I have further relevant information about the contents of the leather-bound book in Smith's library, the second specification leaves me, in the most obvious sense, none the wiser as to what type of figure the envelope contains." (1982, p.62)*

So far so good, but the ambiguity is in what counts as a *revealing* piece of information. Foster claims that the phenomenal contents of experiences are always transparent in his sense, and in this case his claim seems to be that, though we may make a mistake about what a given phenomenal content represents (if it is a component in a hallucination), or about what predicates should apply to it (if we do not know how to use the predicates in question correctly), we can never make a mistake about which phenomenal contents we are apprehending. If there is a patch of some particular colour in my visual field, I cannot be mistaken *that* it is in my visual field, and I cannot be mistaken about what it is like, whether or not it was caused by something that was like it in any way at all.

This leads Foster to the assertion that transparent knowledge of the intrinsic content properties of the mental is unproblematic. It's unclear whether this is true in all cases - we might, for example, ask if there is transparent intrinsic content to a propositional belief - but this is a much smaller question than the problem posed by the intrinsic content of the physical world.

The problem for the realist is how we are supposed to attain any knowledge at all of the intrinsic content properties of physical objects. Under realism, the processes by which we normally expect to learn about the world are all ultimately causal - they answer the question 'What does this thing do?', not 'What is this thing like?'.

To expand a little, we learn about the world in two main ways; through sense-perception and through scientific experiment. Sense-perception is an obviously causal process; (physical) phenomena cause responses in our sensory organs which then pass sense data along to the mind, also - it is normally assumed - causally. It is trivial to point out

that we only apprehend the results of scientific experiments by sense-perception. The more important point in this regard is that scientific experiments study the world by observing the results of controlled interactions between objects; intrinsic content properties cannot be tested in this way because (at least as Foster defines the term) they can have no causal effect on other objects.

So, the realist has no empirical access to any intrinsic content property of any (physical, fundamental) object. Foster, being a strict empiricist, never seriously considers that epistemic access to intrinsic contents might be possible by other means, but I cannot see any method of gaining such access which would satisfy the majority of contemporary realists. By contrast, at least if there is a robust distinction between what a thing is like and what it does - its intrinsic content and its causal properties - the idealist has no such problem.

Ramseyan Humility

While Foster's writing on the intrinsic content problem sometimes sounds as if it exists in a vacuum, much of what we've just looked at should be familiar. It's effectively the same problem discussed by David Lewis in his influential paper 'Ramseyan Humility' (2009), though Lewis' approach to the problem and the terms in which he formulates it are a little different.

For Lewis, the problem arises when we consider how to (completely and correctly) describe the world. Lewis starts by assuming there is some 'final' scientific theory, a theory which describes and can predict, at least probabilistically, everything that exists and that will happen. He does not assert that we can expect to ever finish discovering the final theory, but only that the universe is so arranged that there *is* such a theory.

So far, so unproblematic. And if there is such a theory, it ought to be possible to state it (again, even if *we* could not hope to do so). Lewis then suggests we 'Ramsify' the statement of this theory, so that it becomes a statement purely about relationships between roles - we replace every term[78] of the theory with an existentially quantified variable, predicated with everything the theory says about the role played by the referent of the original term.

Lewis takes this step as unproblematic, and it probably is. It's worth remembering, though, that it may turn out that such thoroughgoing Ramsification is impossible. It might

[78] Or at least, every term that appears to refer to what anything is like, perhaps even every non-logical term.

be that we *do* need to say something about some *occupant* of one of the roles set out by the Ramsified theory to completely describe reality. This would count as a blow against Lewis, though not against the intrinsic content problem in general, and particularly not against the argument based on it that we are interested in[79].

We can see Lewis' version of the problem coming: as he puts it, "It is one thing to know that a role is occupied, another thing to know what occupies it." (2009, p.204) Furthermore, he outlines an extremely plausible case that for at least some roles, we have no idea what occupies them. That is, where both of a pair of roles are captured by any particular kind of predicate (eg. Binary monadic, binary dyadic, scalar monadic etc.), a given property in the reality described by the theory could be the occupant of either interchangeably[80]. We have no way of knowing *which* candidate for the occupancy of certain structurally similar roles actually *is* the occupant of the role.

Because Lewis' discussion (and much of the work which has followed it) is limited to what are effectively scientific theories dealing purely in terms of how different properties relate to one another, it is a simplification or perhaps a sub-case of Foster's broader intrinsic content problem. But it does allow us to indicate more clearly exactly the direction Foster is moving in (and the direction we shall follow).

The argument we shall discuss in the next chapter turns on a structural ambiguity very much like that posited by Ramseyan humility. Given a structure which putatively captures the relationships between all fundamental properties but which can be 'filled out' (instantiated) by the actual properties in multiple different ways, we shall ask how, if at all, it is possible to tell which instantiation of the structure actually obtains. The answer will require us, in essence, to turn to mental intrinsic natures (phenomenal contents) and thus introduce a form of idealism.

But that is a task for the next chapter. First we must examine whether there is any way we might solve the problem of humility, and the literature on the topic is rich with approaches to this question. I can give only a shallow survey of a few options here - the topic would be large enough for a book all to itself.

---

[79] In fact, Foster's point is exactly this; without knowing something about the intrinsic natures of some of these role-occupiers, we leave relevant questions open about the world, but this is a deep point we shall have to return to later.

[80] The inverted colour spectrum (between worlds rather than between minds) is an example of this - provided there is one unique phenomenal colour for each colour role, it does not matter which phenomenal shade plays which role. It makes no difference to us, provided we can all agree which role is instantiated at any given point.

Responses to Humility

In his paper, 'Ramsification and Knowledge-*Which*' (2008), D.T. Locke gives the following, slightly tongue-in-cheek overview of the responses to Lewis:

> *"[P]hilosophers tend to have one of three reactions to Lewis' thesis.*
>
> *Reaction 1: Ramseyan Humility is just a form of traditional skepticism.*
>
> *Reaction 2: So what?*
>
> *Reaction 3: Huh?! [a.k.a. 'the perplexed stare' (not to be confused with the incredulous stare)]" (2008, p.2)*

The thrust of Locke's argument is in favour of reaction 2; he argues that while it is true that we have no epistemic access to intrinsic contents[81], at least of physical objects, we should not be worried by this fact because nothing which affects our lives could possibly *turn* on which intrinsic nature goes where in the Ramsified theory. I shall show in the next chapter that this is false - specifically, the question of whether idealism or realism is true of a particular portion of the world will turn on this issue.

Before that, though, we should take a brief look at the arguments underlying reaction 1. Schaffer (2005) gives perhaps the most direct version of this response. Schaffer actually takes himself to be defending 'quidditism' - the view that properties have intrinsic natures which are not captured by anything in the Ramsey sentence describing Lewis' final scientific theory - against a set of objections brought by various exponents of forms of causal structuralism (a topic to which we shall turn shortly).

Let's simplify the theory Schaffer is defending and put it in our terms. Our interest is in a theory which permits the existence of possible worlds which are structurally identical to ours but which differ from it purely in terms of what things occupy the structure. These worlds might be worlds where some structure-occupants from our world are interchanged (as, for example, in a world with an inverted phenomenal colour spectrum, where longer-wavelength light is blue and shorter-wavelength light is red), or they might be worlds whose structures are (partly) filled in by occupants which do not feature in our world at all. We might state this theory thus:

---

[81] The response to Lewis has tended to prefer the term 'quiddities' (specifically the intrinsic natures of properties) because Lewis' discussion is limited to ambiguity between property-structures. For now, I will follow Foster in using the broader term.

Fosteran Humility: there is some possible world, structurally identical to ours, which differs from it purely in virtue of what intrinsic natures appear at what nodes of the structure.

The sceptical argument can then be unpacked something like this:

1. If Fosteran humility is true, then we do not know which intrinsic natures exist.
2. We do know which intrinsic natures exist.
3. So Fosteran humility must be false.

The jump from Fosteran humility as formulated above to premise 1 is that, given Fosteran humility, even if all possible alternate worlds actually only have interchanged intrinsic natures (so that they still fill out the Ramsified structure with the same *set* of intrinsic natures), we have no way of knowing this, and no way of ruling out the possibility of a world which does have some intrinsic nature somewhere in its structure which doesn't feature in our world[82]. This is not problematic from our point of view.

Premise 2 is, in a sense, question-begging, in that it is a condition of exactly the thing we are arguing against. If we know which intrinsic natures exist, then we must either know that there are or that there are no physical intrinsic natures. But anyone propounding this argument is perhaps entitled to say that premise 2 is the intuitively obvious position, or true for other reasons, or at very least that we haven't done nearly enough work to show it to be false.

Schaffer's strategy is to argue (in favour of premise 2) that humility is true only for the kind of strict definitions of 'knowledge' which have long since fallen out of fashion because of their inability to handle traditional sceptical scenarios like Descartes' demon. Schaffer himself is interested in defending a metaphysics which endorses intrinsic natures (specifically, quiddities), but which *also* allows knowledge of them. So he wants to say that, by any *reasonable* standard of knowledge - standards by which, unlike the strict standard, ordinary knowledge is possible - we *can* know about intrinsic natures[83].

By a roundabout sort of route, we could construe Foster as ending up at something like this position, certainly in his earlier work, but Foster's approach has metaphysical

---

[82] Lewis discusses this point in detail in his original paper, at pp.212-4

[83] It may be, though, that the knowledge Schaffer has in mind is not the transparent knowledge Foster is interested in. If this is the case, then Schaffer is not at all in opposition to Foster, but his arguments remain interesting if we take them in the stronger sense.

consequences - consequences we will actually want to endorse - which Schaffer's lacks. Here is Schaffer's position:

> *"[O]ne can extend (or perhaps parody) the argument for quiddistic scepticism to an argument for external world scepticism, as follows:*
>
> *(25) If there is a difference between appearance and reality, then we do not know what is real;*
>
> *(26) But we do know what is real;*
>
> *(27) Therefore: there is no difference between appearance and reality." (2005, p.20)*

(26) here should be taken as implicitly offering some standard of knowledge - inference to the best explanation (abductionism), say, or contextualism, according to which we do know what is real; a standard of knowledge which renders irrelevant the traditional sceptical scenarios. And, Schaffer says, this same standard will do for intrinsic natures/quiddities.

Abductionism makes a good example. We can take Schaffer as saying 'the explanation that the intrinsic natures match up with the real world is better - a more effective or efficient explanation - than alternatives'. It doesn't posit any dubious switching around (as for example in the inverted colour spectrum case), to be concealed by the fixed natures of roles. It can be stated more simply.

But is this true? D.T. Locke disagrees. He gives a deeply technical discussion of the general distinction between traditional scepticism and 'quiddistic' scepticism and of the problem with the abductionist response in particular (2008, pp.3-6 and 2009, pp.229-236). It's easier to state it in Lewisian terms first; Locke argues that, given that the Ramsified theory suffices to describe reality[84], it must capture all our explanatory resources as regards reality. So merely quiddistic variations between worlds *make no difference*.

The point generalises well beyond abductionism, and something very like it will be crucial in our discussion of Foster's arguments in the next chapter. Intrinsic natures are, by definition, inert (except in virtue of their contingent link to roles in the Ramsey sentence). If knowledge of them is possible at all, it can only be by some utterly direct form of acquaintance; everything else, as Foster pointed out, is a matter of some empirical process

---

[84] At least, in terms of all possible observations - which is something Lewis specifies, and correctly identifies as unproblematic.

and thus non-intrinsic. This makes it very difficult to explain how we can refer to particular intrinsic natures, and thus how we can actually *ask the question* 'are these intrinsic natures interchanged?'.

If intrinsic natures can be known at all, it must be by direct acquaintance. And it's not implausible to suggest that direct acquaintance is necessarily privileged (in precisely the way defined in chapter 1). When we discuss the phenomenal contents of experiences, for example, we do so by reference to the public objects responsible for those experiences. If I ask you to characterise purple, purely in terms of *what it is like in itself*, you will not be able to answer[85].

Direct acquaintance generally, I think, cannot be expected to give us non-private-language means of *referring* to what something is like (this is the problem presented by Feigl for subjectivity-based definitions of the mental, which I discussed in some detail in chapter 1). But then, how do we phrase the claim 'I know that such-and-such an intrinsic nature fills this role in the Ramsified theory'? With what should we replace the 'such-and-such'?

More technically, we can say that our ability to transparently (in Foster's sense) describe intrinsic natures is limited to 'non-rigid' descriptions; descriptions which change from possible world to possible world. For example, we can point to some intrinsic nature in our world and say '*this* is the intrinsic nature which fills this role in our world'. But just as well, our counterparts in some world where a different intrinsic nature fills the role can say the same thing, and be just as right, without us being able to explain how the statements differ. If we could explain how the statements differed, we would be making an observation[86] about how they differed, but that observation would, ex hypothesi, have to be expressible in the Ramsified theory, so the distinction would no longer be purely one of intrinsic natures.

While it's true that we must be able to indexically identify intrinsic natures in a rigid way, by pointing to some intrinsic content and saying 'everything (across all possible worlds) that is like *that*', this will not be a transparent specification. It will not be a specification which enables us to say anything about the content in question, because everything that can be said about it must be captured by the Ramsey sentence of Lewis'

---

[85] Even synaesthetic gobbledegook like 'if it were a sound, it would be a fanfare' can be no help here - that's not what it's like in itself, it's what it *would be like* if it were some other self.

[86] The idiom 'make an observation' is here not misplaced or reified; when one 'makes an observation', one *states* an observation one has made, and that is exactly what is required to distinguish the our-world statement from the their-world statement.

final theory. This is the key difference between quiddistic scepticism and ordinary scepticism - all standards of knowledge require us to be able to say things about the things we know about, but anything that is expressible must be expressed in the Ramsey sentence and thus no standard of knowledge, however generous, can get beyond the Ramsey sentence.

The Perplexed Stare

I take the preceding to show that, if there are intrinsic natures or quiddities, knowledge of them is impossible in just the way Foster envisages. Before we go on to discuss the most substantive of the alternative positions, it's worth taking a moment to examine how (D.T.) Locke deals with the problem. Above, I pointed out that 'merely quiddistic variations between worlds *make no difference'*, and this is a key part of Locke's response.

Locke takes us through a detour into what it means to ask a particular kind of question - questions concerning 'knowledge-which'; that is, knowledge about which of a set of propositions is true. What Fosteran humility prevents us knowing, putatively, is *which* combination of intrinsic natures fills out the Ramsified structure in the actual world (because Fosteran humility says that there are multiple, indiscernible possibilities).[87] So what does it *mean* to ask which filling-out of the structure obtains?

Locke argues, following Aloni (2005) among others, that there is a contextual element to this question. More precisely he argues that 'knowledge-which' questions always have a contextual element. Aloni's example is the difference between knowing the name and knowing the face of a person; each of these can be the answer to the question 'which (of these options) is such-and-such a person'. If the question is 'Which of these men is the president of Mali?', followed by a list of names, one knows the answer if one knows the name of the president of Mali, but not if one only knows what he looks like. If the question is followed by a collection of photographs, the reverse is true.

If 'knowledge-which' questions always involve a context, then something must specify that context. Locke advocates van Rooy's (2003) account of how this context is fixed. This requires the correct answer to be fixed by a 'contextually salient decision problem'. Simplifying, because van Rooy's account is technically complex, this means that

[87] Actually, Fosteran humility has another important consequence - that we cannot form any positive, transparent conception of what the physical world is like. For present purposes, though, this is not particularly relevant.

the person asking the question must be trying to reach a decision about which of his possible actions he should actually perform. The answer must be capable, in essence, of *making a difference* to someone.

And Locke is right, I think, to argue that quiddistic knowledge is not (in general) capable of satisfying van Rooy's criterion. If it is possible to completely Ramsify a complete description of the world, then any question that makes a difference to the behaviour of some element of the world must be captured in the Ramsified description, and therefore involve no quiddistic knowledge. So, by van Rooy's criterion, we can't give which-quiddity questions a context, and they come out meaningless.

I am not convinced by Locke's argument, though. There are three problems with the progression he outlines. We can ask whether he (with Aloni) is right that knowledge-which questions always require a context (and what is meant in this case by 'context'). We can ask whether van Rooy's criterion is the correct way to provide that context. And we can ask whether there really is absolutely nothing that turns on quiddistic knowledge.

If I say that the meaning of some sentence is contextual, we would ordinarily take that to mean it changes depending on the manner and situation in which the sentence is uttered. A simple example is indexicality; the statement 'I am male' is true when I utter it, but not when my sister does, because the meaning (or at least the part of it which is constituted by its reference) of 'I' is contextual - it depends on the identity of the utterer. Similarly, the meaning of 'More than half the occupants of this room are male' is contextual, because the meaning of 'this room' depends on the location of the utterer, and the truth of the sentence also depends on how many men and women are actually in the room picked out.

In what way does a 'knowledge-which' question get its meaning from its context? 'Knowledge-which' questions are of (or, plausibly, can all be restated in) the form 'Which of these options is X?'; the context gives meaning to the phrase 'these options'. Now, quite trivially, this context can for some questions be made explicit in such a way that the question ceases to be context-dependent. For example: 'Which of Richard Nixon and Lyndon Johnson was more recently President of the United States?'

And, though less elegantly, any 'knowledge-which' question with a finite answer set can be rephrased in this way. If I am at an embassy dinner in honour of the visiting President of Mali, my surreptitious question 'Which of the people at this dinner is the President?' can be rephrased with a conjunction of the names (or descriptions, or images, as preferred) of everyone at the dinner in place of the phrase 'the people at this dinner'.

This is a step in the right direction as far as the defender of meaningful humility is concerned, but we face a thornier problem yet. If Lewis is right - and I think he is - to assert that there are 'alien' quiddities (quiddities which exist in some possible worlds, but not ours), then we cannot finitely specify the options for which set of quiddities - or intrinsic natures - occupy which nodes in the Ramsified world-description. There may still *be* a finite specification, but we cannot make it and, honestly, we have no reason to expect that, if there are possible non-actual quiddities, they exist only in limited supply.

And this may seem as though it supports Locke's position that no meaning can be given to the question of which set of intrinsic natures actually fills out the Ramsified description. I disagree, though I accept that my argument on this point is weaker than the points made above. I think it is possible to replace the context-sensitive part of any 'knowledge-which' question with a (potentially very complex) structural description that all answers must satisfy.

So our earlier question, 'Which of the people at this dinner is the President?', would become something like 'Which of the participants in the formal dinner taking place at the embassy on such-and-such-a-date is the President?'. To check if someone is a possible answer to the which-question, we simply ask 'Is/was so-and-so at the dinner?' - and this is not a which-question.

And, by definition, we have a structure which is fit to check whether a set of intrinsic natures is fitted to be an answer to the question 'which set of intrinsic natures actually satisfies the Ramsified description?' - any set of intrinsic natures which *could* satisfy the Ramsified description is eligible. So while we can't finitely specify the options among *which* we are trying to select, we can give an infinite specification.

This may not be the most convincing line of argument, but it is only the first of our complaints against Locke. The second is that van Rooy's criterion seems to me to be far too strict as a criterion of *meaning* for which-questions. It's a good criterion of utility, for sure, but while I accept that a which-question which doesn't meet van Rooy's standard will be fundamentally idle, unlikely to have a determinate correct answer, I don't think it's right to say all such questions are meaningless.

For example, consider the question 'Which of Superman and Batman would win if they had a fight?'. This is a textbook example of an idle question; the characters in question are entirely fictional and to my knowledge the question has never been decisively

answered in any canonical work of fiction[88], so we're unlikely to get a satisfactory answer one way or the other. But we know what would constitute an answer, and we can contextually specify what it would mean to know that answer, by specifying who counts as an authority in the matter. And we can imagine that a question of this type might have no behavioural consequences - might resolve no decision problems of van Rooy's type.

Is the question meaningless? Of course not. One can ask it, and expect other people to have some idea what one is getting at. More work is needed - work specifying the context - to make it precise enough to answer, but that work need not involve decision problems for anyone the question is asked of. We might change the question to 'According to this comic in which they fight, which of Batman and Superman wins?'. This is a legitimate way of answering Aloni's problem without invoking van Rooy's answer.

So it's entirely possible that 'Which set of intrinsic natures actually satisfies the Ramsified description?' is a meaningful, but idle, which-question. Can we go one step further, and argue that in fact the question is *not* entirely idle? I think we can, though not here. Precisely this investigation is at the heart of Foster's arguments against realism, which we will discuss in the next chapter. In short, we will discover that whether (a certain kind of) idealism is true or false depends on whether Fosteran humility is true or false[89].

Causal Structuralism

This, as far as I am concerned, puts an end to Locke's 'perplexed stare' response. Fosteran Humility - the intrinsic content problem - is an inescapable limit on our knowledge. But this is only the case if there actually *are* (as Locke and Schaffer both accept) intrinsic natures or quiddities. What happens if one attempts to deny that such things

---

[88] Here, context is undeniably significant - we could specify any of several ways of generating a 'canonical' answer; we could ask the owners of the Superman and Batman intellectual properties (or the creators of the characters, or look at the works published by those owners or creators), or we could take a survey of all stories featuring both characters, or we could poll their fans, or we could attempt to work out in detail what these characters' hypothetical capabilities would be and simulate a representative sample of their possible conflicts. I'm sure most such approaches have been tried by someone on the internet already.

[89] Technically, we are interested in whether Fosteran humility is true *of the physical world*, the structure described by our best theories of physical science (which are all approaches to Lewis' final theory). If it is, then there are many different ways that the structure could be 'filled out' with intrinsic contents, and it is this on which Foster's argument will turn. The *mental* structure, which Foster takes to be a completely separate thing from the physical structure - a position which requires some defence - is filled out transparently by phenomenal contents. As already stated, I don't intend to go too deeply into the debate about phenomenal contents and knowledge by acquaintance. Lewis and Locke both raise the possibility (Lewis calls it a form of panpsychism; Locke thinks of it as direct realism) but reject it for complex or unstated reasons.

exist? Foster discusses this possibility in detail as the 'powers thesis'. It involves taking the physical world to consist purely of dispositional and causal powers – dispositions and powers to affect other dispositions and powers. As Foster puts it,

> *"The situation will be like that of a supposed language in which the only source of meaning lies in the meaning relations in which terms and expressions stand to one another. Just as a language can only have meaning if some of its terms or expressions make direct semantic contact with things that lie outside it, so the total system of physical powers can only have content if at least some of its components are powers to affect something other than the behaviour of items of power." (2008, pp.68-69)*

Foster's point here is an obvious one - it seems as though, if one tries to specify everything in terms of relations among causal powers, the final specification will be circular. A power, quite plausibly, is defined in terms of *manifestation conditions* - circumstances that cause it to act on something - and *effects* - what it actually does. But if all manifestation conditions are systems of powers and all effects are behaviours of powers, there doesn't seem to be anything to give the whole system an ontological foundation at all.

Objections of this kind, which I shall collectively call the *circularity objection*[90], have appeared consistently through the history of causal structuralism. While I do not wish to advocate causal structuralism in the slightest, I have to say that, as a *metaphysical* theory, I am not convinced it faces a problem in the circularity objection. More on this below, but first we should consider three options that Foster places before his structuralist.

Foster's options all focus on answering the circularity objection by suggesting ways to 'fill out' the structure. The first is to say that while there are no actual intrinsic content properties, it is possible that there might be at some point. This approach claims that the system of powers can be filled out by reference to these possibilities, by saying that each power-item *could* possess this or that intrinsic content, were such a thing to come into being.

---

[90] This is a loose grouping. Bird (2007b) identifies numerous versions of the circularity objection, including Foster's own (grouped with one owed to Swinburne), and a version a little closer to the one expressed here, attributed to Howard Robinson and E.J. Lowe. We shall return to Bird's answer to the circularity objection in the next section; I feel it is strong enough that the opponent of causal structuralism is better served by looking elsewhere for argument.

This is so manifestly unsatisfactory as to be almost incoherent. In the first place, it offers no solution to the problem of epistemic access to intrinsic contents, so even if it is possible that some intrinsic content might come into being, the realist would be no better off than he is now. Presumably Foster's hypothetical realist intends that the possible intrinsic content should be some sort of mind-independent stuff, but it is unclear how we might allow as possible the existence of such while still acknowledging that even should it exist we could have no clear idea of what it is or might be like.

The second escape Foster offers the realist is to say that some of the powers in the system are powers to affect human experience - whether in content or structure is irrelevant. As we have seen, phenomenal contents make good candidates for intrinsic content properties because they are immediately and transparently available to us - we can quite naturally apprehend their full natures, not just their associated causal roles. Unfortunately for the realist, there's a reason phenomenal intrinsic content properties were discussed as an advantage of idealism; phenomenal contents are - must be, by their very phenomenality - intrinsically mental. So on this view, the only intrinsic content of the physical world - the system of causal powers - becomes mental.

In Foster's view, this is surrendering not just the field but also the castle to the idealist. His goal is to prove that there is an ontological dependence of the physical world on the human mental realm. If, as he assumes, the metaphysical circularity he identifies is problematic because of a lack of ontological foundations, and this deficit is filled with mental content properties, then the physical structure turns out to depend on its mental content.

The realist's best hope on this second strategy might be to plead that what is proposed is actually a form of dualism. This is a possible claim because the strategy separates the realm of human experience from the physical system; it cannot reduce the human mental realm to part of the system of powers or it would be no step forward at all. However, if the realist wishes to claim this view as a form of dualism, he must acknowledge that it is a physically non-realist form; nothing in his system justifies any claim to the existence of mind-independent physical stuff[91]. There is a pseudo-physical system separate from but dependent on the realm of human mentality, but it is pseudo-physical only in that

[91] The resources of this system are the human mental realm and a set of powers to affect it or other powers. There's nothing here that has a non-mental intrinsic nature, and no space to insert a non-mental intrinsic nature for anything without running afoul of the same basic epistemic problem.

it seems mind-independent to perceiving humans. In fact, the idealist system Foster eventually lays out is very similar to this position.

The third solution is the only one which seems both viable and acceptable to the realist. It is to say that some of the power-items are powers to affect the geometrical structure of space. However viable it is, though, Foster is far from clear precisely what he means by 'powers to affect the geometrical structure of space', above and beyond his example of the general theory of relativity. On this model, spacetime is bent around objects of mass in such a way that any object travelling in a straight line through the affected area will have its path bent (though the effect is only minimal except around bodies on the planetary scale).

This hangs all the metaphysical work that intrinsic content properties must do about the shoulders of the property of mass. Prima facie, this is quite sensible, since mass is the measure of the quantity of (what's usually assumed to be) physical stuff which constitutes an object. If any physical property could play the relevant role, mass seems the likeliest candidate. Except, of course, that to have an effect on the geometrical structure of space is to be a causal property, so it cannot be mass itself which is the source of intrinsic content for the powers structure[92]. Instead, the intrinsic content involved is the intrinsic content of space itself, and the powers system is anchored by reference to that.

There is another problem here, too, though philosophers will probably be quite evenly split over whether or not it matters. This is the problem that this approach seems to be committed to a Newtonian model of space (certainly, Foster's discussion following on from this point explicitly treats spatial points as part of the fundamental ontology). Whether or not a Newtonian ontology is problematic, it is a necessary consequence of this spatial version of the powers view.

On the contrasting Leibnizian position, space exists purely as a matter of relations among fundamental particles - whether these are monads, in line with Leibniz, or something else. But if space (and especially its structure) is just a set of relations among powers, then it collapses into the circularity objection, because the circularity objection just is the problem that there is nothing to give ontological foundations to powers when all we have is relations among them.

---

[92] If further demonstration were needed, the Large Hadron Collider's main purpose is to search for the Higgs Boson, the hypothetical particle responsible for mass - and as previously discussed, anything which science can reveal cannot be an intrinsic content property.

Against the Circularity Objection

All of this, of course, only matters if there is any force to the circularity objection, and as suggested above, I am not sure that there is. We saw at the end of chapter 2 that Bradley was probably wrong to claim that the interdependence of properties and relations presented an ontological problem. Something similar seems to be true of the powers thesis, at least when construed as a metaphysical theory.

That is to say, the circularity is unproblematic because it is not an ontological regress. It does not require one to step down through an infinite sequence of ontological layers to reach the fundamental level - all the dependencies among powers, putatively, are at the same ontological level.

The same is not true of the powers thesis in the form it was originally developed, which was as an epistemological theory. Construed epistemologically, the powers thesis makes it impossible to pin down the nature of any power well enough to use it as a point of reference to get a hand on any other - any attempt to do so would lead to an endless chase around in circles through the network of powers.

Fortunately for scientific investigation, there is a simple way to deal with this problem; we can note that certain physical powers seem to be associated with the production of mental intrinsic contents - phenomenal properties - and get a handle on the structure that way. This may be the essence of what Foster has in mind with his second option, above, that the powers structure can draw its content from its relation to human experience (it's certainly true that Foster equivocates between treating the powers thesis as epistemological and metaphysical in his discussions).

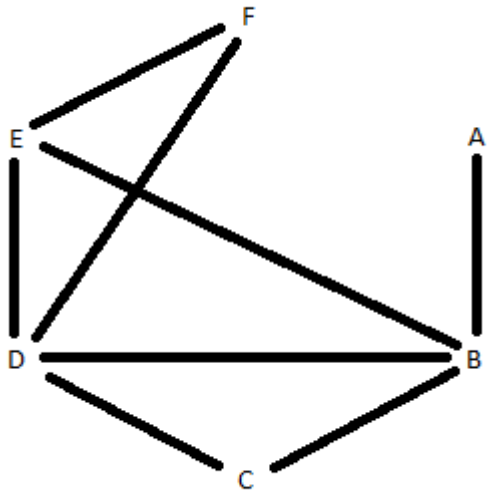
It is possible that, from a purely scientific point of view, this can be achieved in a metaphysically neutral way (if we take, as I believe we should, scientific theories as making statements relating primarily to the apparent world *as if its constituents were metaphysically fundamental*, whether or not they are), but even if it cannot, this is not an argument I intend to pursue against realism or structuralism.

The complaint I made against Bradley can serve as a response to the circularity objection, but I think there is perhaps a stronger answer. This is provided in its most sophisticated form by Bird (2007a), building on work by Dipert (1997). They argue that there are certain special circumstances in which a structure *does* give us enough information to fix the identities of causal powers; specifically, when that structure is (or can be represented by) an *asymmetric graph*. An asymmetric graph is a graph each of whose nodes can be uniquely identified purely in terms of its place in the structure.

Taking a graph to be a set of nodes and a specific set of relations between them, we can offer the following example of a six-node asymmetric graph:

- Node A: Stands in only one relation (to another node).
- Node B: Stands in four relations, one of which is to the single-relation node.
- Node C: Stands in two relations, both to four-relation nodes.
- Node D: Stands in four relations, none of which is to the single-relation node.
- Node E: Stands in three relations.
- Node F: Stands in two relations, one to a three-relation node and the other to a four-relation node.

Visually, the graph can be drawn (most simply) like this:



But the important point is those descriptions. Every node is described purely in terms of how it fits into the structure, and each has a unique description[93]. If there is some way of interpreting a graph of this kind as a model of the world (as consisting purely of causal powers and relations), I think there are reasonable grounds for rejecting the circularity objection. It may well not be the case that the powers thesis is inherently circular, and even if it is, the problem I already raised for Bradley suggests that this may well not be a vicious circularity. This is not to say, though, that the graph-theoretic approach is without its problems.

[93] The fact that this makes the ontological fundament *purely* relational - diametrically contra Bradley - is not necessarily a problem (and in fact is what Dipert sets out to prove).

These problems all relate to exactly what the nodes in a given graph are supposed to be. Dipert remains agnostic, suggesting in the end (as a 'wild and possibly irresponsible speculation') that "[W]e should perhaps consider seriously the possibility that something like the pan-psychism of Spinoza, Leibniz or Peirce is true, and that vertices are pure feelings (Peircean 'firstnesses'), constituting a distinct thought or object only when connected to other such entities." (1997, p.358)

This, fairly clearly, is a position that would fit under Foster's second response to the powers theorist, as given in the previous section. In Foster's terms, it would certainly constitute a form of idealism, and it would also assure the possibility we are seeking to establish, namely that the world could have been such that different 'pure feelings' occupied different nodes. As such, Dipert's view is not one we need trouble ourselves with further.

Bird (2007a, ch.2), on the other hand, is a more stringent causal structuralist. For him, the nodes of the structure are causal powers, defined by their manifestation conditions (the powers that cause them to have an effect) and their effects (on other powers when manifest). This is structuralism *par excellence*, and explicitly requires that the essence of things be exhausted by their causal powers.

Oderberg (2011) offers a rather charming critique of Bird's position. An asymmetric graph can be rendered symmetric by the addition or subtraction of a number of relations, which can be achieved by the addition or subtraction of nodes. A key part of Bird's and Dipert's claims is that only an asymmetric graph suffices to constitute a world. So, it seems, it must be logically impossible for some - and perhaps any - causal powers (on Bird's interpretation) to go out of existence, on pain of the rest of the world following them.

This is a brief and perhaps facile summary of Oderberg's argument. His point is that the graph-theoretic causal structuralist seems to be committed to the position that existence is not contingent. Thus stated, the view seems unusual, but it may be that causal structuralists (who will ordinarily be quite deflationist about possibility, since by definition they will be causal essentialists and deny Lewis' Ramseyan humility) are willing to bite the bullet on it.

The problem I want to consider is one arising from the attempt to be simultaneously a causal structuralist and a physical realist. On our terms, to be a physical realist is (among other things) to hold that the things that are ontologically fundamental are all spatial, so if a realist is also a (Bird-style) causal structuralist, he must hold that the powers which occupy nodes in the structure have clearly-defined spatial locations.

Power-Instances and Location

It may help at this point to distinguish powers in their general form (eg. the power of gravity) from what I will refer to as their *instances* and their *manifestations*. An *instance* of a power is a disposition; for example, the gravity of Jupiter is the disposition of things (relatively) near Jupiter to be drawn towards it. A *manifestation* of a power is the actual effect of that disposition on some particular object, so the orbits of Jupiter's moons are each separate manifestations of its power of gravity.

We cannot really get ahold of Foster's problem without deciding whether the nodes of Bird's graph theory are powers, instances or manifestations. We can, I think, rule out powers; while an epistemic structuralist such as Russell or Carnap may want to think of our best scientific theories as being structures of abstract powers (captured, perhaps, by an equation such as that of the Standard Model of quantum physics), an ontological structuralist such as Bird is committed to the claim that the concrete world, the world we are in, is composed of elements of such a structure.

The version of causal structuralism that takes the graph-nodes to be manifestations is a strange one; it is deflationist in quite a Humean way about cause while trying to preserve a purely causal ontology. As a Humean deflationist myself, this is the only version of structuralism I find even remotely plausible (though I do not think it can be fully developed), but I do not believe there are any actual causal structuralists who take this view. As such, at least for now, we can discard it.

This leaves us with the power-instance version of the theory. The problem I want to raise with this view is a dilemma of sorts concerning how to do justice to the dispositionality of instances. An instance I of a power P, we can say, characterises a region of space R such that entities within R are disposed to behave P-ly. But the thoroughgoing structuralist is committed to a Leibnizian account of space and time (because Newtonian space, it is supposed, is not purely relational[94]), so R must in some way reduce to relations among the entities that occupy it.

On a Leibnizian theory of space, R reduces *ontologically* to some or all of the relations which characterise the node of the world graph occupied by I. So I's relations to

---

[94] It might be possible to give a quasi-Newtonian account of space that was purely structural, by claiming that the points of space exist but get their identities purely as nodes of a spatial structure. This, though, is a hard position to reconcile with Bird's, since there do not seem to be grounds for claiming that the structure of space itself is an *asymmetric* graph, and so the nodes of the spatial structure do not meet this strict condition of identity. We will return to the question of a more explicitly Newtonian partial structuralism later.

other nodes must suffice to constitutively sustain R. This can be achieved in two ways: either we construe all I's relations as causal, and say that some or all of them also have spatial implications[95], or we say that some of I's relations are causal and some spatial.

For reasons that cannot be quickly summarised, we cannot yet deal with the second of these options. This will have to wait until the end of the next chapter, where we shall see that it does not present a problem for the argument of Foster's that we will discuss. The first option would present a problem for that argument, except that it cannot be made plausible.

The key point about this option, which I will call *purely* causal structuralism, is that it allows nothing except what is sustained by causal relations. And the problem is, roughly, that causal relations could only be sufficient to sustain a physical space if they were already spatial - that space must come first.

I think it is obvious that, to be capable of playing a constitutive role in a Leibnizian model of space, a causal relation must have what I'll call a *spatial implication*. For example, the force of gravity between two objects (of a certain mass) is inversely proportional to the square of the distance between them; if you know the masses of both objects and the forces acting on them, you can calculate the distance between them. The spatial distance is implied by the masses and forces.

Masses and forces, quite plausibly, are purely causal properties, and can thus be captured by a purely causal structure. Any causal structure which furnishes objects with mass and forces of gravitational attraction, then, ought to suffice to sustain spatial relations between them. This, I take it, is how the purely causal version of causal structuralism is supposed to preserve the spatiality of the physical world.

Causal relations which have spatial implications, then, like the gravitational attraction between two masses, can be called space-sustaining relations. The problem with these space-sustaining relations is in *why* they have spatial implications. Normally, when two objects interact causally, we take them to be doing so *in virtue of* their spatial relationship. When magnets repel each other - exert a force upon one another - it is because they have been brought close together. When a comet or space probe 'slingshots' around a planet, the gravitational effect occurs because of the proximity of the two objects.

At very least, the structuralist seems to have got this backwards; he says that the comet and the planet are close together because they interact gravitationally, and the

---

[95] Or are necessarily one-to-one correlated with spatial relations - the key point being that on this model there can be no spatial relation where there is no causal relation.

magnets are close together because they are repelling. But perhaps the structuralist will insist this is unproblematic, that we should not complain if the ontological fundament turns out to be stranger than we expect - after all, idealists draw some pretty strange conclusions too.

The problem runs deeper than this, though. A space-sustaining relation is a causal relation with a spatial implication, but this spatial implication must be *specific*: at a given time, it must have a fixed magnitude (must set a fixed value on the spatial distance it sustains). The magnitude is supposed to be fixed by the magnitudes of the various causal properties in the purely causal structure, but these magnitudes are (in part) conditioned by spatial distances.

If two objects, each massing a million kilos, attract each other with a force of one Newton, this means they are approximately 67 metres apart. But while we can put the question of explaining mass aside[96], it seems to me implausible to say that the force between (acting on) the objects is not conditioned in some sense by the distance, rather than vice versa. The explanation for the magnitude of the force just is the proximity of or distance between the objects, and I cannot see another explanation that would serve.

The purely causal structuralist can say that the magnitudes of his relations occur as brute facts, but from a scientific point of view this is frustratingly limiting. It seems like it would make the mathematical formulas of scientific law coincidental - lacking an explanatory principle that would guarantee them - since it denies that other facts have any bearing on the behaviour of powers at all. Powers behave as they behave, and we have only luck to thank for the fact that they do so in some physical-world-sustaining way. And there may be a more serious problem yet.

This is the problem of non-manifesting powers. I take it that, ordinarily, a causal structuralist will want to support the idea that powers - dispositions - continue to exist while they are not acting on anything, with some sort of principle or condition to guarantee *when* they will manifest, and *how*. A magnet floating out in the deep void between galaxies somewhere will not attract or repel any other object, but we ordinarily take ourselves to be able to say what it *would do* to an object that it *came* to affect.

But we do this in spatial terms. The lonely magnet will attract or repel other magnetic objects which come close enough to it. The problem is that on purely causal structuralism, 'coming close enough' means 'being affected by'. So another magnetic object

---

[96] This is because, at time of writing, this is a very open question from a scientific point of view.

will be attracted to or repelled by our lonely magnet when it is affected by it - and this seems to be a trivial statement. It certainly doesn't seem to express a meaningful *manifestation condition* for the magnetic power of the lonely magnet. It doesn't actually explain whether or not the power will manifest in any given instance.

Putting this more formally, we want to be able to say that a given power P will produce an effect E on an object O if O is within the region R[97]. On a Leibnizian account of space, regions reduce to spatial relations among their occupants, and on purely causal structuralism, spatial relations reduce to causal relations. So R must be sustained by some causal relations. But what causal relations? P is currently not manifest. It stands in no causal relations.

It is true that P stands in many relations of *potential* cause, i.e. that it could have a causal effect on many other objects. But none of these relations have a spatial implication - they have no fixed magnitudes from which a spatial distance could be calculated. It's simply possible that P could affect them.

Perhaps a particularly hard-headed purely causal structuralist could argue that space is constituted by those powers which are currently manifesting, and that non-manifesting powers are derivative of the web of currently manifesting powers. This, though, has the bizarre consequence that every time a new power begins or ceases to manifest, the structure (and thus, under purely causal structuralism, the identity) of physical space changes. I cannot imagine there are many philosophers who would prefer this solution to abandoning causal structuralism.

[97] Actually, in truly space-sustaining cases, we want to say that the *strength* of E is related to position within R, but this does not lower the bar at all for the structuralist.

5. The Experiential World

Moving Beyond the Spatial Strategy

In chapter 3, we saw that there is a distinctive and problematic limit on what we can know about the world. The problem that arose there was that both Kant's and Bradley's 'spatial' strategies for idealism depended on assumptions about the external, mind-independent world (if there is such a thing) which could not be checked. For example, while Kant arguably could be considered to have proved that the appearance of spatiality that we experience is a creation of our 'pure intuitions' - and thus our minds - there is nothing in his argument to justify a leap to the conclusion that the realm that *causes* those experiences is non-spatial. It could, quite independently of our experiences, be spatial, either in a way that was or was not reflected in the arrangement of our experiences.

I do not intend to challenge this epistemic limit any further. There may be arguments which allow us to engage with the fundamental world, but I do not have them. Instead, I want to turn now to an argument about the ontological status and constitution of the world a step 'above' the fundament, which I think we can deal with more effectively.

The world[98] in question is the world of common-sense, macroscopic objects - tables, chairs, people and so on. These are all objects which, according to our best scientific theories, are *constituted* by fundamental particles (whatever those particles are thought to be), but which, according to our ordinary way of thinking, are something over and above the particles that constitute them.

I suppose it might be possible to deny the claim that a table is more than the sum of its atoms, but I think this would be a mistake (and almost certainly not one anyone has made). I do not intend to claim that there is *very much* more to a table than the sum of its atoms, but there is something more to a table than to some other groups of atoms of equal size. Say a table contains 10^{27} atoms; a table is something more than a pile of 10^{27} atoms of carbon, or an arbitrarily-selected 10^{27}-atom-sized lump within the rock of Mount Everest.

The point does not need belabouring; all that matters for now is that we acknowledge a distinction, even if only a slim one, between the fundamental world and the world it constitutes. To deny this, in essence, is to stray into a particularly strong form of naive realism, which takes the world of tables and chairs to *just be* the fundamental

---

[98] Perhaps 'world' is too strong a term - 'realm' or 'ontological level' might be clearer, but this use of 'world' is consistent with Foster's and thus worth sticking with for a discussion of his work.

world[99], and I shall discuss this position in the next chapter. For now, I want to demonstrate that the logical distinction just identified has an unexpected ontological consequence.

We must start by getting completely clear about what we are discussing. Earlier, I described the constituted world as a world one step 'above' the fundament, but this was on the assumption that there is only one stage of constitution - that tables and chairs are directly constituted by fundamental particles. On our best contemporary scientific theories, however, one could argue that tables and chairs are constructed from bits of wood, metal and plastic, which are constituted by molecules, which are constituted by atoms and so on. For this reason, it helps to be more specific.

The world I am interested in is the world of tables and chairs, whether or not it is only one ontological step above the fundament. It is the world that we ordinarily think of our experiences as bringing us into contact with. Our experiences do not (directly) tell us about molecules and atoms, but about tables and chairs. For this reason, I shall call the world I am interested in the *experiential world*, E.

That said, I do not mean, by 'experiential world', the world *of* experiences (the world, or representation of a world, composed of my externally-perceptive mental states, if there is such a thing). To do so would make any argument for idealism trivial, since we would by definition be talking about something mental. It might be most useful to think of the world we are interested in as the one which *directly causes* our experiences.

The question is not whether or not the experiential world is physical - I think it quite obviously is. It is spatial, its elements are quantifiable in terms of units which seem to be real (in the sense defined in chapter 1), and many facts about it are not privileged to any particular subject. But the experiential world is, by definition, not ontologically fundamental - it is constituted. The question, then, is whether there is a mental element to this process of constitution or not.

A scientific physicalist will want to claim that the process by which E emerges from the fundamental world (call it F) involves no facts about human (or any other) mentality - that it is captured entirely by objective quantifiable and spatial relations among fundamental particles. We can call this position reductive or scientific physicalism; I take it to be, as much as there is such a thing, the normal way of thinking about the physical world.

---

[99] To do the opposite - deny the distinction between the fundamental and constituted worlds while insisting that it is the world of atoms which is fundamental - would be analogous to eliminative materialism in the theory of mind. One would have to deny that the constituted world and its objects existed at all, which I take to be gibberish.

What I will argue, following Foster, is that reductive physicalism is false, and that there is an essentially mental element to the constitution of E by F. To put it another way, the existence of E depends, ontologically, on some mental fact or facts. Specifically, Foster's argument makes the claim that the physical space of E (the space inhabited by all elements of E) depends on the relation of F to human experience, and that if F were differently-related to human experience, E would possess a different physical space.

In this sense, we could consider Foster's argument a version of the spatial strategy, but it avoids the problem encountered in chapter 3 in virtue of its reduced ambition. It does not attempt to prove that the fundamental world is non-spatial, only that the spatiality of the world that we (directly) experience is a mental construct. Foster points out that it is most natural to interpret the phrase '*the* physical world' as meaning what I have called E; there may be many other worlds which are physical, but we have little reason to be interested in them. His question is whether *the* physical world, the one we inhabit in the most immediate sense, depends ontologically on the mental.

It is worth noting that the main thrust of Foster's argument (which I shall refer to henceforth as the 'functional arrangement argument') is against a Newtonian reductive physicalism which treats space as substantial. For most of what follows I shall use 'space' in a Newtonian sense, though I will show, eventually, that something like Foster's conclusion stands even if we switch to a Leibnizian model.

Nomological Organisation

Foster actually gives two quite distinct versions of the functional arrangement argument, one in his 1982 *The Case for Idealism* and the second in the 2008 *A World For Us*. Both are immensely complicated and written in a style that, while effective, takes some getting used to. I will cover each in turn; the earlier version has the strength of being more general, but is perhaps less convincing than the latter, which has the weakness of presupposing mental realism[100].

In *The Case for Idealism* (hence CFI), Foster sets out the argument in this way;

> *"(1) The geometrical structure of P [the physical space associated with E] is essentially linked with the nomological organisation of P... (2)*

[100] In his *The Immaterial Self* (1991), Foster gives an extensive and compelling defence of mental realism - the thesis that the human mental realm is not reducible to the physical - which there is not space to do justice to here.

> *P possesses its geometrical structure essentially … (3) No component of the physical world [E] is ultimate unless P is." (1982 p. 128)*

The concept of the 'nomological organisation' of a physical space may be an unfamiliar one, but Foster is using it in a very precise sense that we must be careful to understand before proceeding. Nomological principles are defined by the Oxford English Dictionary as those which "resemble laws, especially those laws of nature which are neither logically necessary nor theoretically explicable, but just are so." (Oxford, 2013)

Nomological principles, then, govern how things behave, but are contingent. It might seem at first glance as if Foster's claim (1) is trivial, since it is only natural to assume that the geometrical structure of a space has some effects on objects moving through it. For example, if P's geometrical structure is uniformly Euclidean, then any P-occupant travelling in a straight line will never cross its own path.

This, however, is not a nomological principle, since it is necessary - it is part of the definition of a Euclidian space that straight lines do not cross themselves. Foster explicitly excludes this kind of principle. His claim (1) is that some elements of the nomological organisation of P which are not a priori deducible from its geometrical structure are nevertheless essential to P's possession of its geometrical structure.

A better example of a nomological law, then, might be that motion is always continuous - that no-one ever suddenly teleports to a point fifty feet ahead of where they were just walking. There is nothing in the geometrical structure of P to explain why this should be. The particular set of nomological laws that Foster will use for his argument, however, are very different. (1) should seem a very contentious claim, and we'll get to his argument for it shortly, but it's worth pausing before we do to understand how the three points given above achieve his aim if accepted as premises.

What isn't obvious from this way of stating the argument is where the mental is supposed to come in[101]. The essential, hidden point is that the component of the nomological organisation of P to which P's geometrical structure will turn out to be essentially linked is the set of laws governing how P-occupants appear, and appear to behave, to us. Foster's argument for (1), then, will be that our experiences of E have a role to play in its constitution, and this will be where we spend most of our time.

---

[101] And, from an exegetical standpoint, it may well be that in CFI Foster was not so clearly addressing himself to this objective as in the later *A World For Us*. It will be much easier to explain the latter, though, once we have a good grasp of the former, whether or not the former succeeds.

Giving a brief summary of how this works is very difficult. Our experiences represent the physical space of E as having a certain geometrical structure, call it G_E. The physical space, P, of E also (putatively) possesses a certain geometrical structure, call it G_P. Logically, it seems possible that these could differ, but what Foster's argument will show is that this is not the case; G_p must be (or must be structurally identical to) G_E.

This might be thought unproblematic (after all, an ordinary indirect realist will assume that, generally, the structure of space as represented to us in experience is represented to us accurately, so of course G_E will match G_p), but Foster's argument goes one step further. It shows that in fact, G_p is the geometrical structure of P *in virtue of the fact* that G_E is the structure represented to us by experience. This is a very strange claim at first glance, but we shall see it defended in due time.

The argument for premise (1), if successful, will therefore show that P, the physical space associated with E, only has its geometrical structure in virtue of some mental (or at least mind-involving) facts; specifically, those facts which capture how P appears to be arranged geometrically. This refutes the broadest forms of physical realism, but leaves two potential slightly weaker forms available. Premise (2) addresses the first of these and premise (3) the second.

The first option would be to accept that G_P is dependent on some mental facts, but insist that P does not possess its structure essentially and thus can still be ontologically independent of the mental. Foster's argument against this is that the identity of a space is wholly constituted by its geometrical structure, so that there is no circumstance in which P could possess a structure other than G_P. If the physical space of E possessed a different structure, it would not be P, but some other space.

The second option would be to acknowledge that Foster's argument shows that physical space is derivative of some mental facts, but to deny that this ontological dependency spreads to the occupants of physical space. This move could be seen as Leibnizian in spirit, since the most obvious way of making it would be to say that physical objects are ontologically basic, but their nomologically-conditioned behaviour and appearances constitutively sustain P.

Given our definition of 'physical', though, it is not clear that this will help the realist at all; either he will have to accept that the ontologically basic objects are not essentially spatial, and thus not essentially physical, or he will have to allow that their space-sustaining behaviours and appearances are essential to them. This latter is a strange position in itself,

but it reintroduces the essential link to experience encountered earlier - they can constitute E only in virtue of fitting themselves to G_E.

These summaries are unavoidably obscure due to the need for brevity; I shall cover each point in far greater detail henceforth.

The Interchange Case

Foster's attempt to prove (1), that the geometrical structure of the physical world is logically inseparable from its nomological organisation, is simultaneously his masterstroke and one of the most baffling parts of his writing. It's not fair to lay all the blame for the bafflement at Foster's feet, since the argument is complicated and subtle beyond almost anything else in my philosophical experience, but it will nevertheless take a great deal of explaining.

Foster asks us to entertain for explanatory purposes a scenario he will eventually prove to be incoherent, a scenario in which the nomological organisation of the physical world makes it seem as if it possessed a geometrical structure other than the structure it actually possesses. He gives an immensely precise technical specification of a simple example of this scenario from CFI p.131-139, but it can be boiled down to the following:

Imagine a universe, U, identical to this one except for the fact that the geometric structure, G, of its physical space, P, is slightly different to that of our universe. Specifically, imagine that G places Oxford in a region of Cambridgeshire which in our universe is occupied by Cambridge and vice versa; Oxford and Cambridge are *interchanged*[102].

However, U-humans are unaware of this strange state of affairs, because the nomological organisation, N, of P includes a complex set of physical laws, I, governing the behaviour of matter and energy crossing the boundaries of the two interchanged regions. Here a little of Foster's technical precision becomes necessary.

Let R_1 and R_2 be spherical regions of P, one encompassing the interchanged part of Cambridge and the other that of Oxford. Let R_1 and R_2 be of equal size, and let them possess an identical internal geometrical structure so that for each physical point x within R_1, there is exactly one corresponding point in R_2, y, spatially related to all other points in R_2 in exactly the same way as x is to all other points in R_1. This is most important at the

[102] This is rather simpler to understand on a Newtonian model of space - in U, Oxford occupies the set of spatial points which in our universe are occupied by Cambridge. The interchange case (or something like it) can be reconstructed in a Leibnizian context, though, as we shall see later.

boundaries between the two interchanged regions and the rest of P, so for the rest of this example, let x and y be at some point on the boundaries of their respective regions.

The I-laws ensure that any physical entity which reaches the boundary of R_1 at x is immediately transferred, with all its physical properties (mass, velocity etc.) unaltered, to y. For example, if x happens to sit on a westbound lane of the A40[103], then as one drives towards Oxford, one will be transported to y just as one reaches x, y being the point in R_2 at which the A40 enters that region. The teleportation which has just taken place will be undetectable to any observer within U, since the same rule affects any sensory stimuli - light, sound etc. - escaping either region, and a similar rule governs each other R_1-R_2 boundary-point-pair.

Overall, then, the interchange is undetectable to anyone inside U. U-humans would all believe that U-Oxford was in U-Oxfordshire and U-Cambridge in U-Cambridgeshire, the same way we believe Oxford to be in Oxfordshire and Cambridge to be in Cambridgeshire. However, the U-humans would be wrong, though they could never know this, nor even seriously entertain the possibility unless N changed so that some or all of the I-laws functioned differently.

Obviously, the scenario of U is thoroughly unlikely, but given the precise definition of 'nomological' Foster is using - not logically necessary or theoretically explicable - there is nothing currently established to make it impossible, which is what Foster requires. He brands the interchange example a case of 'reciprocal topographical deviance', but the important point is that it constitutes an example of *nomological* deviance.

Foster gives an exhaustive technical specification of nomological deviance, but we can summarise at least a little for present purposes. Nomological deviance occurs when the way in which P appears to its occupants to be arranged (due to nomological principles) differs from its actual structure[104]. Such a deviation could only be detected by an observer whose powers of observation transcend the nomological laws of their universe; no-one else would have any grounds for claiming there was a difference between the two. We shall come to this presently.

---

[103] Technically, x is a point inside the interchanged part of Cambridge (just), so it would be more correct to say that x *faces onto* the A40, and furthermore that the transfer or teleportation occurs not as you reach x, but *just before you would* reach x.

[104] Henceforth, I shall refer to the way a space appears to be arranged as its *functional* arrangement (or structure, or geometry) and the underlying arrangement as the *intrinsic* arrangement. This is slightly anachronistic, since these terms are only used by Foster in relation to the later *A World for Us* version of the argument, but he uses them there as I have used them here.

'Correct' Geometries

We can characterise the physical space of U as possessing two separate geometric structures, the intrinsic G_1 (according to which Oxford is in Cambridgeshire[105]) and the functional G_2 (according to which Oxford is in Oxfordshire). It should be obvious from the preceding section that G_1 and G_2 are just two of a possibly infinite set of geometrical structures that could apply to the physical space of U (or any other universe). If nomological deviance is logically possible, then any physical space of any universe *could have* a whole range of different geometrical structures, and yet still appear the same to its occupants.

At this point, an eager realist might say, 'So what? Yes, there are lots of weird deviant structures that a physical space *could have*, but to suggest that this universe has a physical space structured in this way offends against Occam's razor. It's far simpler to accept that our physical space is non-deviant.' This misses the point in several ways, some of which we won't get to until we consider the later version of the functional arrangements argument.

The point, ultimately, is that deviance is a relationship between geometrical structures. It involves one geometrical structure being incorrect *relative to* another, and thus requires that there be some reason for holding some geometrical structure to be *correct*. In the interchange case, G_2 is correct for the nomological arrangement, but incorrect for the intrinsic arrangement (i.e. incorrect from the point of view of G_1).

Foster's question is what it means to say that a geometrical structure is correct. We have thus far characterised G_2 as the 'deviant' structure (i.e. the one that is incorrect), because it differs from G_1, which we have specified as being the geometrical structure that P 'actually has'. But why should we accept this as a criterion of correctness? If we stipulate that G_2 is 'correct' (say, because it is the one we actually have to deal with on a day-to-day basis), then G_1 becomes deviant and incorrect.

The point generalises, since every geometrical structure by definition is deviant relative to any other (because two separate geometrical structures which are geometrically identical are simply two instances of the same structure). No *geometrical* fact about any of these structures can pick it out as more significant than its fellows.

---

[105] Technically, because G_1 is a geometrical structure, it would be more precise to put this in terms of the geometrical relation of R_1 to the rest of U. Identifying R_1 with Oxford introduces non-geometrical information to our specification, and we shall see that this would be counter to the thrust of Foster's argument.

We can shed some light on the problem by examining the one clear association we do know about between a geometrical structure and a component of U; the association of G_2 and the nomological arrangement. Specifically, we know G_2 *is* the geometrical structure which characterises the nomological arrangement of P, the physical space of U. We can take ourselves to have defined G_2 thus:

Let R_1 be the region containing U-Oxford, R_2 the region containing U-Cambridge, R_3 the region containing U-Oxfordshire, and R_4 the region containing U-Cambridgeshire. G_2 is the geometrical structure such that R_1 is inside R_3 and R_2 is inside R_4. G_1, by contrast, places R_1 inside R_4 and R_2 inside R_3. Since we know from experience that U-Oxford appears in all respects to be in U-Oxfordshire, and similarly for U-Cambridge, G_2 characterises the functional arrangement of the physical space of U.

The key point is that, if we are to hold some geometric structure up as 'correct', we ought to have good reason for doing so. What we are looking for is the correct geometrical structure of the physical space of E, the world we most directly inhabit. And it seems obvious to me that E's geometric structure must be the structure of its functional arrangement; remember that there is no way for a U-occupant to tell that his universe is deviant, so the deviance which actually occurs in that universe cannot be a feature of E by definition.

This is what makes the link between a physical space's possession of its geometric structure and its nomological arrangement essential. The geometric structure of a physical space *just is* the one which captures its nomological arrangement, even if there is some other arrangement of its fundamental entities. And the link makes the dependence on human (or at least conscious) experience clear, too: it is only the geometric structure which captures how things appear to us[106] which can serve.

Geometric Structures and Indeterminacy

The issue is not yet completely settled, however; Foster discusses two possible ways in which the realist might argue against the nomological thesis. He takes himself to have refuted both, though in the case of the first he is perhaps guilty of requiring lower standards of himself than of his opponents. The first objection is that under the nomological thesis as just outlined, it is entirely possible (if unlikely) that some physical

[106] Or at least, how things appear to us most consistently, in some way that allows us to sort out and exclude illusions and hallucinations.

space might have a nomological organisation which, by virtue of being 'thin'[107] or inconsistent, is equally suggestive of each of two conflicting geometrical structures.

The problem here is that, according to the nomological thesis, this scenario involves a necessary contradiction. The nomological thesis entails that each of the two suggested structures is 'correct' for physical space, but, purely in virtue of the fact that they are different structures, there must be some geometrical proposition that they disagree over the truth of. So it seems that the nomological thesis will commit us to both the truth and falsehood of some geometrical proposition in this case - and we would normally take this, by reductio ad absurdum, to refute the nomological thesis.

Unusually, Foster does not offer a thoroughgoing example of such a space, but I shall endeavour to illustrate, because I think the objection is rather stronger than Foster's treatment would suggest. Let P be the physical space of a universe U, N be its nomological organisation, and G_1 and G_2 two competing geometrical structures, both candidates to be the structure of P.

U is similar to our universe, except that there is a tiny[108] region, R_1, of P for which N includes a special set of laws, the T-laws. These function much like the I-laws in the interchange case example as far as R_1 is concerned; any particle crossing the boundary of R_1 is translated to the counterpart point on the boundary of another region, R_2. However, the T-laws are not interchange laws, because particles crossing the boundary of R_2 enter R_2 quite normally; they are not sent to R_1.

So, particles never end up inside R_1, but only ever in R_2. The essential point for our purposes is what the T-laws say about particles *leaving* R_2, which is that they have an exactly 50% chance of ending up in the region around R_1 (R_3) or the region around R_2 (R_4). To put it another way, a particle leaving R_2 will *seem to have been* in R_3 half the time, and in R_4 the rest of the time.

We can imagine that this state of affairs arises because of some complicated experiment in teleportation. Thus, R_3 and R_4 can be different laboratories, one at Cambridge University and the other at Oxford University, and R_1 and R_2 the focal points of the equipment set up therein. Now let G_1 be the geometrical structure which matches our

---

[107] This is Foster's term, and I am not completely clear what he means by it, though I assume it is something along the lines of 'not involving very many nomological rules which have geometric implications'.

[108] To be specific, R_1 is of such small size that only a single fundamental particle can fit inside it at once. The reason for this will become obvious shortly.

own universe and places R_2 inside R_4, and G_2 be some other structure which identifies R_2 with R_1 and thus places it inside R_3.

The question is whether or not there's any way of telling which of $G_1$and G_2 is instantiated by P. We can put this to the test by sending some particles at R_1 and R_2 and seeing what happens to them, but first we need to define how to interpret our experimental answers.

This can be done as follows. For any regions R_n and R_o, R_n is inside R_o iff:

1. It is possible for a particle K travelling inside R_o to cross any point on the border of R_n without passing through any intermediate region (i.e. any region that is not a sub-region of R_o).
2. Any particle K travelling inside R_n will, if it travels far enough without changing direction[109], emerge back into R_o.

Let's start by imagining that a scientist in the Oxford laboratory sends a stream of particles towards R_1. We can assume that the equipment setups in both the Oxford and Cambridge laboratories include detectors near R_1 and R_2 respectively to check where particles going through the experiment end up.

We can test point 1, above, by moving the Oxford particle emitter around R_1 while keeping it pointed *at* R_1. And what we will find is that every particle emitted behaves as if R_2 is in R_3; that is, there is no point on the boundary of R_2 which cannot be reached from R_3[110]. Each particle in the stream ends up at the point on the boundary of R_2 which is T-law-correlated with the point at which it left R_3, and R_3 allows us to direct the stream of particles at any point on the (internal) boundary of R_3.

As far as point 2 goes, each particle in the Oxford stream reaches the boundary of R_1 and teleports to R_2. It then crosses the boundary of R_2 and emerges either in R_3 (Oxford) or R_4 (Cambridge), and does so exactly half the time in each. Thus, fifty percent of particles behave as if $R_2$is in R_3 and the other fifty percent behave as if $R_2$is in R_4.

---

[109] These two conditions - travelling far enough and not changing direction - are rendered insignificant in our example because R_1 is so small. Because only a single particle can fit into it at once, any motion in any direction will cause the particle to cross the boundary of R_1. The only problematic example would be if a particle managed to stop exactly inside R_2, and even then, we can assume or stipulate that the next particle to cross the boundary of either R_1 or R_2 would knock it out again. Either way, there would be a negligible effect on the pattern of our statistics.

[110] Assuming, of course, that the teleportation field doesn't involve travel through some alternate, intermediate space, which is at least logically possible, but can be barred by stipulation.

From the point of view of the Oxford stream, then, all the point 1 evidence puts R_2 in R_3, as does half the point 2 evidence. Points 1 and 2 should have equal weight, so we can add these proportions simply, and say that three quarters of the Oxford evidence puts R_2 in R_3, while one quarter of it puts R_2 in R_4. But the Oxford evidence is itself only half the evidence[111], since we should also test how matters stand in Cambridge.

Let's repeat the Oxford experiment in the Cambridge lab, then. When we point the Cambridge beam of particles at R_2 we find, quite naturally, that as far as point 1 is concerned, R_2 is in R_4; every point on the boundary of R_2 can be reached from with R_4 (remember, there's no teleportation field affecting particles entering R_2 from R_4). And the test of point 2 will produce exactly the same evidence as was produced in Oxford; half the evidence puts R_2 in R_3, and the other half puts R_2 in R_4.

So, three quarters of the Cambridge evidence puts R_2 in R_4, and one quarter puts R_2 in R_3. Sum this together with the Oxford evidence, and we get half of our total evidence putting R_2 in R_3 and half of it putting R_2 in R_4. If R_2 is in R_4, G_1 is correct; if R_2 is in R_3, G_2 is correct, and this is the only way of deciding which of the two structures is correct[112]. Thus, we (or U-humans) have no way of telling which of G_1 and G_2 is correct.

So, we have a possible if improbable (though no more so than the interchange case) world in which there are two equally good candidates for the geometrical structure of physical space and none better. Are we then to say that P possesses no correct geometrical structure? Or that P is not a physical space, despite all appearances to the contrary?

Foster is dismissive of this concern; he says it is enough for his nomological thesis to allow that P has an *indeterminate* geometrical structure, provided we accept that P can only count as a real physical space as long as each of its candidate geometries is 3-dimensional and the discrepancy between its candidate geometries is not too great.

It is this last condition where the problem rests; Foster does not (and almost certainly could not) specify the precise degree of similarity required between candidate geometries for P to still be a real physical space. Now, leaving questions of degree open on the assumption that *some* degree could be adjudged the correct one is a long-standing philosophical tradition, and often entirely defensible.

---

[111] This can be achieved in two ways; either by analysing the numbers purely as proportions, or by stipulating that both labs perform the experiment the same number of times. Either will do.

[112] There is a legitimate question over what happens to R_1, or rather to the gap left behind in the middle of R_4 by G_2's identification of R_2 with R_1. I believe this could be resolved, but only by making the overall scenario exponentially more complicated. I have skipped an attempt to do so for brevity's sake.

Unfortunately, I am not sure this is such a case. Foster's nomological deviance argument is intended to establish that the nomological arrangement of a physical space is essential to that space's possession of its geometrical structure, but his overall strategy, at least in CFI, also requires that a physical space possesses its geometrical structure essentially. It seems strange to me to suggest that a physical space - or anything else for that matter - could essentially possess an indeterminate[113] geometrical structure, so Foster seems committed on this point to saying that in fact the above example involves no real physical space at all.

Nor can Foster claim that the indeterminacy is only epistemic - that there is some correct geometry which we (or U-humans) lack the information to identify. It may be that U possesses some intrinsic arrangement which has a definite geometrical structure, but that is irrelevant for our purposes. The nomological thesis (or rather, the way Foster's defence of it works) entails that the geometrical structure of P is constituted by its nomological organisation; P has the geometrical structure that the nomological organisation conditions. And in this example, the nomological organisation either conditions two contradictory structures, or none at all.

But if U is a universe without a physical space, because its nomological organisation fails to provide an adequate geometrical structure for one, then 'having a nomological organisation which provides an adequate geometrical structure' becomes a necessary condition for a thing's counting as a physical space. And this makes it trivial that the nomological organisation of a physical space is essential to its possession of its geometrical structure. I am unsure if this triviality is out-and-out harmful to Foster's argument, but it does open him up, in a very roundabout way, to the allegation of defining his conclusion into truth. This is, I think, the greatest weakness of the CFI version of the functional arrangement argument.

The Transcendental Thesis

The other objection Foster considers, one to which he gives considerably greater attention, he titles the 'transcendental thesis'. This is a view which claims that physical space has an intrinsic geometry discernible from some special perspective unaffected or at

[113] It may seem that the idealist can argue that the indeterminacy is only epistemic and not metaphysical, but this transgresses the point of the nomological thesis. The point there was that the only way in which a geometrical structure can *be* the structure of a physical space is because it is the way that physical space seems to be structured to us (or the occupants of the relevant universe). Thus, as far as geometrical structures go, epistemic indeterminacy *is* metaphysical indeterminacy.

least undeceived by the nomological organisation. The obvious way to fill this theory out is to say that the special perspective is (a not necessarily actual) God's. According to the transcendental thesis, then, the geometrical structure of physical space is that of its intrinsic arrangement, as discernible by superhuman observers, and the nomological arrangement is a deception or illusion in some way.

So far so good, but the transcendental thesis doesn't really help us humble humans; it leaves us unable to identify the correct geometrical structure of the space we inhabit, and, indeed, leaves us without any grounds for certainty that physical space even *has* an intrinsic structure. Foster offers one suggestion (I can see no alternatives compatible with the transcendental thesis) according to which the transcendental perspective is sensory, and the physical world is (or is ontologically derivative of) the sense-field belonging to the holder of that perspective.

This is, of course, pretty much Berkeley's system, and it does solve the problems just given. It is possible for an appropriately divine-sized sense-field to possess a geometrical structure of the right sort, and, arguably, we can trust that God will ensure that a physical world is constituted and sustained for our benefit. Interestingly, this picture of the transcendental sense-field thesis has some striking similarities with the idealist model Foster proposes at the end of *A World for Us*, but I do not intend to discuss that model in detail.

For now, Foster has objections to bring against this version of the transcendental thesis, and by extension, any other that might be dreamed up. He gives two problematic examples; the first involves a case where instead of a single sense-field providing the intrinsic geometry, this work is done by a pair of linked sense-fields, each responsible for half the structure, and the second involves a scenario where, despite the physical world being functionally three-dimensional, the sense-field underpinning it has only two dimensions.

Obviously in the first case, we do not have an adequate geometric structure for physical space; the two sense-fields cannot combine effectively for this end[114]. It is tempting to say, however, that the case is incorrectly formulated because a perspective

[114] This should be obvious if we take seriously the definition of mental given in chapter 1 (I take it as uncontroversial that sense-fields are mental; this is certainly what Foster had in mind). We saw there that the appearance of spatiality in our sense-fields cannot itself be a genuine spatiality because there is no way to precisely determine the spatial distances involved without invoking a non-mental standard of comparison. Something similar goes for the present case; there is no way to give geometrical relations between points in the first sense-field and those in the second without invoking something over and above them, and there is no obvious candidate for what this could be.

split in this way is insufficiently transcendental - that is, it's insufficiently general. The essence of the transcendental thesis is that there is a single, unifying, correct perspective on the geometrical structure of physical space, that propositions of physical geometry are true insofar as they match the intrinsic geometry of whatever underlies it. The holder of the transcendental thesis may therefore feel comfortable with the claim that in this example there is no physical space, because there is no single, unified underlying geometry. Either way, the second case is the more problematic.

The problem here is supposed to be that only a three-dimensional space can qualify as *physical* space, so if the underlying space is not three-dimensional then it is not the one we are interested in. This would mean that whether or not there is a 'correct' geometry from a transcendental perspective (i.e. whether or not U has an intrinsic geometrical structure that is knowable to some possible God) is irrelevant.

But it is not clear that this assumption, that physical space must necessarily be three-dimensional, is well-motivated. It is true that the physical space of E is three-dimensional, but the transcendental theorist wants us to ignore the physical space of E - his claim is that what we should be interested in is the (putatively) physical space of F, underlying E. Understanding this, Foster's assertion that only a three-dimensional geometrical structure will serve begins to look a bit like begging the question, particularly since there are theories within physics which suggest that the universe might indeed be intrinsically two-dimensional (more on these shortly).

Overall, then, Foster's objections to the transcendental thesis aren't nearly as conclusive as he seems to think they are. At very least, a rather more substantial discussion would be required to settle the issue. However, few realists are likely to find the transcendental thesis attractive - it is altogether too Berkeleyan. It may remain a possibility, but it is a slim one, and I am happy to follow Foster in moving on.

Beyond the Nomological Thesis

Let us therefore allow Foster his nomological thesis; the nomological arrangement of the physical word is essential to its possession of its geometrical structure (and more specifically, has a part to play in *determining* that geometrical structure). This completes the first step of Foster's core argument, leaving us to deal with the much less difficult second and third steps. The second step is to show that physical space possesses its geometrical structure essentially and the third is to show that if physical space is not ontologically fundamental, no part of the physical world can be.

To a certain extent, we have been proceeding on the assumption of the second step, because it is simpler and far more intuitive to assume that physical space is logically inseparable from its geometry, particularly given how used we are to a physically realist mode of thinking. However, the nomological thesis radically changes the conditions which make this common position attractive. Specifically, if the nomological organisation of the physical world is essential to its possession of its geometry, but itself contingent, it becomes attractive to the realist to say that the physical world only possesses its geometry contingently, thus preserving the ontological fundamentality of the physical world.

Think of it this way: the realist wants to preserve the fundamentality of the physical world. If the geometrical structure of physical space depends on its (contingent) nomological organisation - i.e. the nomological thesis is true - and the geometrical structure of physical space is essential to it, then the contingency of the nomological organisation is 'passed along' to physical space itself. That is, physical space turns out to be ontologically dependent on something which is not fundamental (the nomological organisation), so by definition it cannot itself be fundamental. But if physical space is essential to the physical world - i.e. no part of the physical world is fundamental unless physical space is - then that dependence gets passed on one further stage and the physical world also becomes ontologically dependent. Given the nomological thesis, the obvious place to break this dependence is at the claim that physical space possesses its geometric structure essentially, if possible.

This is what Foster calls the ‘realist nomological thesis’ (RNT). Its fundamental claim is that, while the *points* of physical space are ontologically fundamental, its structure - the set of distance relations among its points - is fixed by nomological principles rather than any essential feature of the fundamental world. If this is coherent at all (and I'm not sure it is), then it can be unpacked in two different ways.

On the first, call it RNT-A, the structure of any given physical space is fixed by its nomological organisation, so that in any possible world with a different nomological organisation, the physical space of that world will have a different structure to the one it possesses in our world. On the second version, RNT-B, however, the structure of physical space in any possible world is determined by the nomological organisation of the *actual* world, the nomological organisation we are familiar with.

RNT-B is actually a weakened form of the transcendental thesis, since it singles out some particular geometry as correct irrespective of whatever nomological organisation applies (in any non-actual possible world). It adheres to the nomological thesis only in that

the structure of physical space *could have been* different if the nomological organisation of the actual world was different.

Foster's description and example, given at CFI pp.163-168, are among the most hopelessly dense passages of his writing. What follows is my attempt to reconstruct his argument using a slightly different example. We must start with a set of spatial points, call it P. These are the points which, in the nomological organisation of the actual world, constitute physical space. The realist nomological thesis says that these are ontologically fundamental entities.

However, it says that the *physical* structure they instantiate is determined by the nomological organisation. As I understand Foster, we need to be able to speak of three different kinds of structure instantiated by these points: there is the physical structure, the nomological structure (which determines the physical structure), and there is the intrinsic structure which the points instantiate irrespective of or prior to their nomological organisation.

The intrinsic structure is necessary because the points are fundamental and necessarily spatial, so they must exist, spatially if not physically, independently of the contingent nomological organisation. The points must all stand in spatial relationships to one another[115] - for each point-pair, there must be some distance that the points are separate by. So we have a set, call it S, of spatial relations between points. It is easiest, then, to think of what varies between structures as the *lengths* of various S-members; the alternative would be to say that each structure involves a different set of relations, but this runs afoul of the intuitive proposition that relations are individuated by their relata. I'm not sure this issue would make more than a terminological difference in this case, either way.

An example will, I hope, make things clearer. Let U be some universe which is nomologically deviant. That is, its nomologically-ordered geometric structure differs from its intrinsic structure (the nomological organisation of U assigns different lengths to some S-members than they, intrinsically, possess). Specifically, U is intrinsically two-dimensional, but nomologically ordered to seem three-dimensional to its occupants.

This is an example that Foster makes frequent use of (we have already seen it briefly in our discussion of the transcendental thesis), and indeed there is a theory in physics, called the holographic principle, which suggests that something like this might be true of

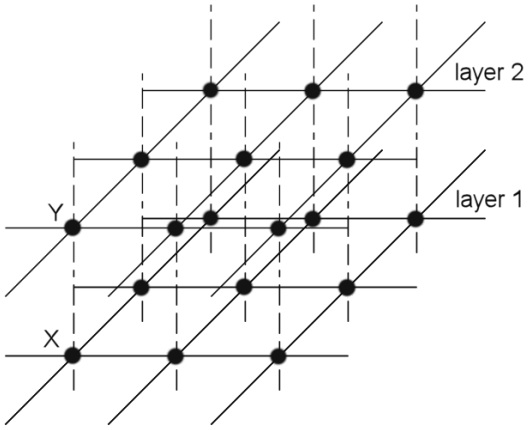
---

[115] At least, so we may stipulate. It's actually conceivable, as discussed above in respect of Foster's arguments against the transcendental thesis, that physical space could be constituted by two or more separate underlying spaces, but as was the case with the transcendental thesis, I'm not sure this would help the realist nomological thesis at all.

the actual universe ('t Hooft 1993). But Foster never spells out exactly how it is supposed to work; furthermore, given the constraints of RNT, the case he has in mind cannot be like that suggested by the holographic principle.

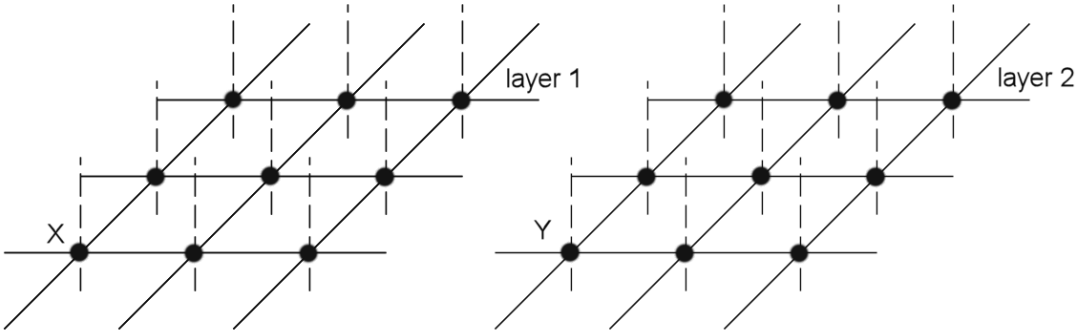
The holographic principle says that the apparent third dimension of physical space *emerges* out of the behaviour of some property or properties spread across a two-dimensional plane. In other words, all the points of the third dimension (those which appear to be 'above' or 'below' the two-dimensional plane) are not themselves fundamental, but constituted. RNT requires that *all* the points of physical space are fundamental.

So, if we are to have an RNT-compatible model of an intrinsically two-dimensional but functionally three-dimensional universe, all the points of the third dimension must somehow be squeezed into the two-dimensional plane. The simplest way of doing this is to imagine that the physical space of U is made up of a series of layers of points in a rigid lattice or grid. Each layer is a single point deep, infinitesimally separated from the layers above and below.



The structure of physical space; each point should be considered to be the minimal possible distance from its closest neighbours.

So each layer by itself would be two-dimensional. Intrinsically, then, we can think of the layers as being laid edge-to-edge rather than one atop the next.



Intrinsically, the layers are separated out and placed next to each other. The dotted lines now represent the nomologically-sustained vertical relations.

The two points X and Y are (physically and nomologically) *vertically adjacent.* A minimal upward motion from X will bring one to Y. This, though, is a nomological principle rather than a feature of the intrinsic arrangement; intrinsically, the distance between X and Y is vast. In the intrinsic arrangement, to reach Y from X, one must cross the entire width of layer 1 to the right of X, then all of layer 2 up until reaching Y. A contingent law allows objects at X to jump to Y without passing through any points in between; a set of laws doing similar work for every other pair of vertically adjacent points is what sustains the three-dimensional physical arrangement[116].

In U, then, the *physical* distance between X and Y is minimal, as small as a distance can be. In physics, the smallest *measurable* distance is called the Planck length, about 1.616×10^{-35}m. For simplicity's sake, let's assume (as some physical theories do) that this is the smallest measurable distance because it is the smallest *possible* distance. So X and Y are physically 1 Planck length apart. Intrinsically, they are the full width of U apart (since they occupy the same point relative to their own layers, everything to the right of X plus everything to the left of Y within those layers adds up to one whole width of the universe).

We can now, finally, clarify the difference between RNT-A and RNT-B. Foster explicitly states that they "agree, of course, on the specification of the physical geometry of [physical space] in the actual world." (1982, p.164) Let us stipulate that U is the actual world - that the world we inhabit actually is nomologically deviant in this radical way. The

[116] The question of whether this framework, in the simple form just stated, does justice to the possibility of diagonal vertical motion, will have to be left for another time, since attending to it now would overcomplicate the example.

two models agree that the actual physical distance between X and Y is the distance actually assigned by the nomological organisation, i.e. 1 Planck length.[117]

What they disagree over is what happens to the physical distance between X and Y in other universes. RNT-A claims that the physical distance between X and Y in other universes is always the distance sustained by the nomological organisation of those universes. So, for example, if some alternative universe U', which possesses the same intrinsic arrangement, happens to be nomologically organized two-dimensionally (i.e. it is *not* nomologically deviant in the way that U is), then the physical distance between X and Y, if there is such a thing, will turn out to be the full width of the universe, rather than 1 Planck length.

Similarly, if we imagine a universe U'', possessing the same intrinsic arrangement but with a three-dimensional nomological arrangement which stacks its layers in a different order, so that layer 1 and layer 2 are not next to each other, X and Y will no longer be vertically adjacent. As such, the functional, nomologically-arranged distance between them must be greater than 1 Planck length. According to RNT-A, X and Y will therefore *not* be *physically* 1 Planck length apart.

RNT-B disagrees. We can speak of the distance, call it d, between X and Y as possessing an *intrinsic* property, of 'being 1 Planck length long according to the nomological organisation of U, the actual universe'. Since the property itself is explicitly universe-bound, d possesses it in all universes. And RNT-B says that the physical distance is set by this intrinsic property, not the 'local' nomological organisation (that is, the nomological organisation which obtains in the universe under question, rather than in the actual universe), so X and Y will always be physically 1 Planck length apart.

Given the preceding, it's not hard to see how this becomes absurd. In U'', where there is a physical space with a radically different structure to our own, and any physical-space-occupant would measure d as being greater than 1 Planck length, RNT-B insists that in fact X and Y are physically 1 Planck length apart. In U', where it is quite natural to speak of there not being any physical space at all, X and Y are still, *physically*, 1 Planck length

---

[117] Actually, there is an alternate reading of RNT-B, hinted at in a note of Foster's at CFI p.298, according to which RNT-B actually holds that, where the universe is nomologically deviant, the physical distances, most correctly, are the intrinsic distances (which brings RNT-B closer to the transcendental thesis). I personally find this version of the theory to be significantly less plausible than the version I present above, since it would make all physical distances in a radically nomologically deviant universe disagree with measurements made by actual physical-space-occupants.

apart, because d still possesses the property of being 1 Planck length long in the nomological organisation of the actual world.

RNT-B might seem more plausible if one insists that the actual world cannot be deviant in this way (so that at least the physical geometry imposed on deviant universes reflects their actual, intrinsic arrangements), but this actually abandons the nomological thesis altogether, since it restores the dependence of physical geometry on an intrinsic arrangement. In fact, this position is nothing more than a return to the transcendental thesis, which we have already dismissed.

This is a very brief summary of Foster's immensely complicated argument, but I believe it serves to deliver the basic point, which is that the spirit of the nomological thesis can be preserved only if the structure of physical space varies with its nomological arrangement. That is, if the same space has a different arrangement in a different universe, then it must have a different structure. This, of course, is exactly what RNT-A says.

But here, too, there is a problem. A physical space must be a particular genuine space, and what makes a given space distinct from any other is in part the geometrical structure it possesses. If S is a space consisting of a (possibly infinite) set of points and possessing a geometrical structure G, then for S to possess some other structure in a possible world is for S to be a different space in that world; the spatial points would have to be in different places (for want of a better expression), and thus would be different points. The points of S are points of S in virtue of being G-structured, so without the instantiation of G by some points, S does not exist.

Foster puts it best:

> *"If* S *is a space and* N *is the network of distance-relations holding between its points, it is logically impossible for those same points to be characterised by a different network of distance-relations. For to be characterised by a different network, they would have to move to different spatial positions in* S*, and thus to become different points. Depending on the nature of* S*, it may be more natural to think of the network of distances as logically flowing from the identities of the points or more natural to think of the identities of the points as logically flowing from the network of distances. But there can be no denying that the two are essentially connected in such a way that we cannot have the same points without the same network." (1982, pp.168-9)*

The point here is that the identities of spatial points are essentially bound up with their spatial relationships to other points. For a given point x, it is part of x's identity that it is (say) five metres from y. The point *ten* metres from y is a different point, call it z, and to change things around so that x is now ten metres from y is to change x into z, not to move x. The point generalises, so that a space with a different structure (a different set of distance relations among its points) is a different space.

RNT-A says that the same physical space can exist across a range of nomologically different worlds. For example, it says that our (presumed non-deviant) world could possess the same physical space as some world in which Oxford and Cambridge were interchanged, in virtue of both worlds having spaces which were constituted from the same set of spatial points. But we have just seen that the same set of spatial points, differently arranged, must constitute a different space, so RNT-A must be false; the same physical space *cannot* exist across differences in nomological arrangement.

Foster considers two replies the RNT-theorist might make to this argument. On the first, the points of physical space get their identities not from their relations to other points of physical space but from the points of some external space that underlies them (Foster's example is a sense-field, as ever, but it is equally possible for it to be an item of a type we are unfamiliar with). Thus, it is claimed, physical space can be constant over variations in its geometric structure because its points are constantly identifiable (they are made independent of or separable from the distance-relations between them).

In effect, what this strategy does is say that it is not the network of *physical* distances between physical points which characterises them, but some network of relationships between the things that underpin physical points. In Foster's sense-field construal, physical points are identified with sense-field points, and characterised by the sense-field distances between the sense-field points they are identified with. It is therefore supposedly possible to change the geometrical relations between physical points without disturbing their identities - the geometrical relations can thus be products of the nomological arrangement.

The problem is that the putative external thing with which physical space is being identified is *not* subject to the nomological thesis; it does not vary its structure with changes in its nomological organisation (in effect, it is equivalent to the intrinsic arrangement discussed above). And this means that *identifying* the points of physical space with it violates the nomological principle, because it prevents the points of physical space varying with their nomological organisation. Given that Foster considers himself to already have proved the nomological thesis, this strategy must fail.

The second objection is neater and handled rather more simply; the objection asks us to think of the geometrical model of gravity proposed by Einstein's general theory of relativity, according to which space itself is bent - that is to say, its geometrical structure is changed - by bodies of mass. Since these bodies can and do move, the geometrical structure of the actual physical space must be variable over time. Foster does not deny this, but instead points out that strictly, under the Einsteinian model, we should factor in time by treating it as a fourth dimension (of spacetime rather than space). Considered as a 4-dimensional environment, physical spacetime is in fact geometrically unchanging[118].

Physical Space and the Physical World

Quite apart from concluding the refutation of RNT, this last argument also establishes the second pillar of Foster's core argument, namely that physical space possesses its geometrical structure essentially. This, combined with the already-explored claim that the nomological arrangement of a physical space is essential to that space's possession of its geometrical structure, gives us the result that every physical space has an essential nomological arrangement. Thus, physical space is contingent and cannot be ontologically ultimate/fundamental.

It's worth taking a moment to unpack this in more detail. Foster's original statement of the argument involved three points: the nomological thesis, the fact that physical space possesses its geometric structure essentially, and the claim - on which we still have yet to touch - that no part of the physical world is fundamental unless physical space is. We can now express the first two of these points more clearly, along with making explicit some steps of the argument which are required for the third of the original premises to become relevant.

The nomological thesis says that the geometry of any candidate for the role of 'being physical space' must vary with the laws which govern the behaviour of the *occupants* of physical space. These laws, though, are clearly contingent - we can imagine many worlds in which different laws to those we observe in our world obtain. So each world that differs in this way from ours has a physical space with a different geometric structure - in fact, because a physical space is intrinsically linked to its geometric structure, each such world has a different physical space.

---

[118] What Foster would say to a presentist, for whom time has no regions besides the present, is not completely clear, but presentism is at best a controversial doctrine on its own merits and a thorough discussion of this alternative is not necessary here.

So physical space exists only contingently (a different physical space could have existed in place of the one that currently exists). It is conditioned by the nomological organisation of physical-space-occupants and thus depends on that organisation. For physical space to be fundamental, this would have to not be the case, so physical space cannot be fundamental.

Putting aside our lingering doubts about the nomological thesis, then, it almost looks as if we have, with Foster, already refuted physical realism. After all, if physical space is not ontologically basic, how can anything else physical be so? The third pillar of Foster's argument is exactly this, that there can be no physical fundamentals without fundamental physical space. His argument is that the scientific, reductive realist position takes common macroscopic physical objects to be derivative of the spatial distribution of intrinsic and causal properties. Make the physical space involved in this constitutive process non-fundamental, and by definition you make the objects produced non-fundamental too.

Leibnizian Physical Realism

The best the realist can offer would be to turn to the Leibnizian account of the nature of space; that is, to say that physical space is derivative of the spatial relations (spatially-related behaviour) of physical objects, objects which are ontologically independent of the space they happen to inhabit. It is entirely possible to reconcile this with our earlier definition of spatiality as spatial-field-membership, because we can construe a spatial field as a collection of spatial relations behaving in a particular, consistent way (and thus as logically dependent on the behaviour of the objects standing in those relations, which can putatively be taken as fundamental).

The Leibnizian model is not uncontroversial, and we have already seen Kant and Bradley criticising it one way or another, but it is by no means out of the question. We should therefore not expect to accept Foster's overall argument without showing that the realist cannot embrace the Leibnizian account of space.

The prima facie problem facing the Leibnizian attempt to defeat Foster's argument is that it appears to make the spatiality of physical objects non-essential to them. That is, the fact that they are spatial objects at all is a matter of how they behave, not what they are[119]. There is a plausible line of defence against this problem according to which we would say

[119] This is because if they were intrinsically spatial, we would have to invoke a spatial field in order to characterise them, but we are bound by the Leibnizian stipulation that a spatial field must be an ontological derivative of its occupants *and not* the other way round; even a circular dependence of field on occupants and occupants on field would contravene the Leibnizian system.

that the spatial behaviour of these objects is essential to their being spatial and thus physical[120], but not essential to the objects themselves.

Foster gives a brilliant but very complicated refutation of this idea in chapter 11 of CFI. I have attempted as best I can to explain that refutation in what follows, but I am sure that in attempting to clarify the argument and make it accessible, I have lost details which may be significant. The argument goes more or less like this:

Let U be some universe consisting of the set of fundamental particles S. According to the Leibnizian strategy under discussion, the physical world of U, P, is fundamental iff the set of physical atoms A *just is* S and A is nomologically organised in a space-sustaining way. So, if P is fundamental, we can expect that physical atoms will move in a consistent, continuous manner (technically, given Leibnizian space, their distance relations will change by smooth transitions), and that the distance relations between them will instantiate a geometric structure which is clearly appropriate to physical space, and so on.

But this is an entirely contingent arrangement, by definition. It could be the case that S-members did not behave in this space-sustaining way. So there is a possible universe U', whose fundamentals are the set S, where S-members are not nomologically organised in a space-sustaining way. For example, in U', the nomological laws could allow some S-members to occasionally move discontinuously[121].

The particular kind of discontinuous movement we are interested in is place-swapping between S-members of some particular intrinsic type. This is because for any two members of S, x and y, such that x and y are of the same intrinsic type, we can assume that they are functionally interchangeable. So if they swap places in such a way that each inherits the properties (mass, velocity etc.) of the other - a scenario whose nomological principles are entirely conceivable - there will be no difference in the way the world seems to be.

This is crucial, because we can stipulate that in U', S-members are correlated to A-members (physical atoms) in such a way as to constitutively sustain a physical world, P'. There is no functional consequence of the envisaged place-swapping, so even if x and y are swapped as envisaged, P' will continue to be sustained. The correlation between x and y

---

[120] As, given the definition of 'physical' from chapter 1, we must take as a basic requirement.

[121] It is much easier to formulate an example if we take it as part of the definition of spatiality that movement must be continuous. While I think this is a plausible assumption, I do not have time to argue for it here, but even if it is incorrect, the Leibnizian account requires that there is some set of behaviours (of objects) which are space-sustaining, and we can imagine that for any universe whose objects behave in the appropriate way, there is some universe containing the same objects (capable of) behaving in a non-space-sustaining way.

and their A-member correlates a and b will swap over (i.e. if x was correlated with a, it will become correlated with b, and vice versa for y), governed by the laws which bring about the nomological arrangement.

This means that in U', A-members (physical atoms) are not identical with their S-correlates (fundamental atoms), because different S-members can be correlated with a given A-member at different times[122]. But now imagine that while the nomological organisation of U' *allows* place-swapping, by coincidence, none ever occurs. U', we can stipulate, has exactly the same fundamental ontology as U (i.e., S), and the members of its fundamental ontology behave in exactly the same way as the members of the fundamental ontology of U.

But if U and U' share a fundamental ontology, and their constituents always behave in the same way, there seems to be no reason to think of them as distinct from one another; they are utterly indiscernible except in respect of an unactualised possibility. Most importantly, the set S as it appears in U is identical with the set S as it appears in U'; each member of S-as-it-is-in-U *just is* the same member of S-as-it-is-in-U'. So x is the same item in U and U'; but in U', we have seen that x cannot be identical with a, its correlated physical atom, because it must be *possible* for x to swap places and correlations with y.

In U, x is identical with a, and in U' it is not, but x in U *is* identical with x in U'. The transitivity of identity thus implies that in U x is not identical with a and in U' it is; a contradiction exists in both universes (since they are, after all, indistinguishable from one another). All this flows from trying to treat the set A of physical atoms as identical with the set S of fundamental atoms, so we can now deny this identification.

But if, even having done away with a fundamental physical space, physical atoms turn out not to be identical with fundamental atoms, then the Leibnizian step has failed outright to help the realist. This step was taken in the hope that some physical objects could be held to be ontologically fundamental even though we have shown that physical space is not; but we have now shown that physical objects cannot be fundamental either way.

Putting this all together, Foster's argument is close to success. To recap, the argument consists of three premises: that the nomological organisation of the physical world is essential to its possession of its geometrical structure, that the physical world

---

[122] If a were numerically identical with x, then it would be necessarily true that if x was interchanged with y, a would follow suit and interchange with b. But there is no logical impossibility in stipulating that this does not occur.

essentially possesses its geometrical structure, and that if the space of the physical world is not ontologically basic, then no part of it can be. The conclusion of this argument is that no part of the physical world is ontologically basic. Foster's arguments for the second and third premises are, I think, solid, but these premises are both *relatively* uncontroversial.

Foster's argument for the first premise - the nomological thesis - unfortunately has a couple of cracks in it. The nomological thesis becomes awkward in situations where two or more geometrical structures are equally close to capturing the nomological organisation of a physical space, because it seems to entail that in these situations physical space will necessarily possess an indeterminate structure. And Foster is not nearly so effective in refuting the alternative transcendental thesis as he thinks he is; his arguments rely on stringent and perhaps unrealistic assumptions about what is required to count as a physical space. Where does this leave his argument? There can be no doubt that the nomological deviance argument is a powerful and problematic one, but Foster's presentation of it in CFI cannot be considered conclusive.

A World For Us

Thus, we must examine the version of the argument, similar in some superficial respects but importantly different, given in *A World for Us* (hence AWFU), which in my opinion fares better, but within a more limited field. If nothing else can be said in favour of the AWFU argument, it does at least make much clearer the involvement of the mental in the constitution of the experiential world, E, by the fundamental, F.

The AWFU version turns on more or less the same example, the interchange case, but characterised in a subtly different way and aiming at an immediate target which is *prima facie* almost diametrically opposite the target of the CFI version. This is because of a fundamental difference in how Foster treats the physical world in AWFU as against CFI, and as a result, far from wanting to show that the interchange case is incoherent (that there could not be an intrinsic arrangement which differed from the functional one), he argues that it is entirely possible and that its possibility entails the fundamental mind-dependence of the physical world.

Following Foster, I'll call the non-mental part of fundamental reality the 'relevant external reality' (RER). The 'relevant' is required because the world we are interested in is the one with which we interact; physical realism as Foster construes it in AWFU consists in the claim that the physical world *just is* the RER. The 'external' should be understood as meaning or at least entailing mind-independence in some ontological sense. The crucial

technical aspect of Foster's use of 'relevant' is that an external reality is only relevant if it is in some sense responsible for our experiences, so the RER is a reality or world on which our experiences depend, but which does not depend on them. The RER is what we have previously been calling the fundamental world, F.

Foster, then, sets out to show that in this picture, the physical world cannot be identified with the RER. The RER has a fairly intuitive composition; time, space, objects and a system of causal laws. It is related (made relevant) to the human mental realm by a second set of laws, called 'link laws', which are fundamentally experiential; they are the laws which connect RER-phenomena to their experiential effects (and mentalistic causes to physical effects). This brings out an important point which was not fully explicit in our exploration of the CFI argument, which is that Foster's argument requires that the human mental realm cannot be ontologically reduced to some part of the RER.

This, of course, is a controversial claim. Foster argues for it at length in his *The Immaterial Self*, but there is not space to do justice to his discussion here. Fortunately, a weaker form of the claim, one which is much easier to defend, will serve. We can allow that it is possible that the mental realm reduces to some (non-mental) element of the RER, provided it does not do so in a way that is mediated by the physical realm (in our terminology, E).

This is because we are not interested (as Foster sometimes seems to be) in the character of the RER; we have seen that there is a very tight limit on what we can infer about it. We are interested in how E emerges from the RER, and if it can be shown that E emerges from the RER only in virtue of some relationship to our experiences, this will entail that our experiences (or some facts about them) are ontologically prior to E. Our mental realm, M, may not be fundamental, but I believe the AWFU version of Foster's argument can show that it does not reduce to some non-mental component of E.

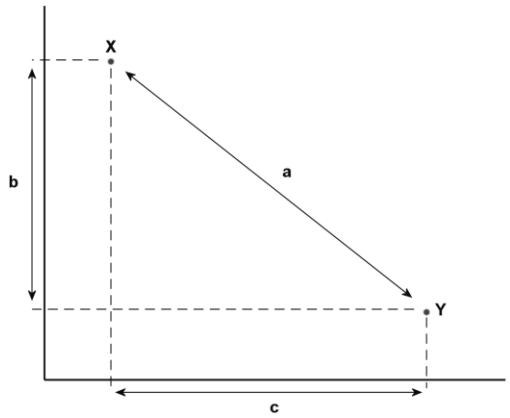
A New Setting for the Interchange Case

The use Foster makes of the interchange case is, given my previous remarks, obvious. Let us assume that in the RER, Oxford is in Cambridgeshire and Cambridge in Oxfordshire, but there is a set of laws, the I-laws, which make things appear, consistently to all human observers and in all sense-modes, as if the opposite is true. Some of the I-laws are causal laws internal to the RER, and others are link laws. The way things stand in the RER, Foster calls their intrinsic arrangement, and the way things appear is the functional arrangement.

Foster's argument, then, is that the physical world must have the form and structure of the functional arrangement, not the intrinsic, and thus that it is not - cannot be - identical with the RER. A simple consideration motivates this claim; as Foster puts it, "The physical world, to qualify as *the* physical world (as the world that forms the target of our ordinary physical beliefs), has to be *our* world, and it can only be our world in the relevant sense, if it is ours *empirically*" (1982, p.138).

Imagine we live in a deviant universe such as the one envisaged. We would have no way of knowing that we did; all evidence would indicate that Oxford is in Oxfordshire and Cambridge in Cambridgeshire. We would regard anyone who insisted otherwise as insane, or at least wrong, even if, from an intrinsic point of view, he was correct.

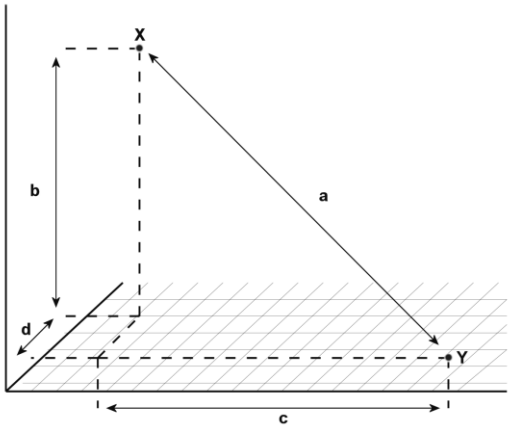
A more extreme example would be the one where the world is intrinsically two-dimensional but still functionally three-dimensional[123]. A little bit of maths will help here. For any pair of objects in a two-dimensional space (points marked on a piece of paper, for example), the distance between them is given by the familiar, two-dimensional form of Pythagoras' theorem, $a^2 = b^2 + c^2$, where a is the distance between the points and b and c are the distances between them in any pair of axes at right angles to each other in the plane.



In two dimensions, the length of a, the distance between X and Y, is given by the Pythagorean formula $a^2 = b^2 + c^2$ or $a = \sqrt{(b^2 + c^2)}$.

In a three-dimensional space, however, three axes are needed to characterise any distance between two points. In this case, the formula for the distance is $a^2 = b^2 + c^c + d^2$.

[123] We have already discussed a version of this example, above, in reference to the realist nomological thesis. There, I mentioned that that version of the example was unusual in that it had to preserve all the points of physical space; this version of the example is, if there is such a thing, a more 'normal' version, and rather closer to the holographic principle in physics.



In three dimensions, the length of a is given by $a^2 = b^2 + c^c + d^2$ *or* $a = \sqrt{(b^2 + c^c + d^2)}$.

Assuming that in the envisaged scenario, two of the three dimensions of the functional arrangement just are dimensions of the intrinsic arrangement, so that b and c are constant between intrinsic and functional arrangements, this means that a will be greater in the functional arrangement than in the intrinsic one, since a third dimension has been added in. Accordingly, the vast majority of statements about distances between two points in the functional arrangement will be intrinsically incorrect.[124]

Worse, we would have no way of knowing of our own ignorance. To insist in spite of this that the intrinsic arrangement, rather than the functional one, is the physical world is to place a very tight limit on the physical facts we can know. Almost everything we take ourselves to know about the position of ordinary physical objects would turn out to be a matter of illusion. For this reason, it seems to me incoherent, or at least hopelessly obstructive, to insist that 'the physical world' is the intrinsic rather than the functional arrangement.

Zero Deviance

Foster's physical world, then, is identical with the functional arrangement of our RER, so long as there is a deviation between the functional and intrinsic arrangements. The remaining question is whether or not we can prove that the point generalises to the (presumed actual) zero-deviance case. Whether or not the physical world can be identified with the RER in any one of thousands of zanily-arranged possible worlds, is it possible to

[124] There are, as mentioned previously, theories of physical science which say that the fundamental reality F is two-dimensional. I do not intend for this argument to be considered a refutation of those theories. Instead, my point (following Foster) is just that when we use the phrase 'the physical world', we cannot be referring to a fundamental world of this type - we must be referring to an intuitively three-dimensional E.

carry out such an identification in *our* world, where, structurally at least, the physical world is indistinguishable from the RER?

It would be a much-attenuated version of realism to retreat to the claim that in the actual world the RER and the physical world are identical, but only contingently upon the functional arrangement of the RER being non-deviant. As Foster puts it, "We surely expect realism, as a *philosophical* thesis, to be a response to our fundamental conception of the physical world, rather than to what we take to be the most plausible account of the specific circumstances of its existence." (2008 p.149) The physical realist wants his theory to be generally, demonstrably true, but in the envisaged scenario, the truth of realism would be a special case, not guaranteed by any fundamental principle[125], and empirically unverifiable. Nevertheless, since it is within striking distance anyway, it is worth spending a moment to close this attenuated realism off as well.

It is at this point that the true strength of the 2-dimensional RER case comes to light (not that the interchange case falls altogether short, but the other makes the key point more obvious). In the final version of that case, the RER and all its internal laws were 2-dimensional, and the third dimension of the functional arrangement was generated wholly by the laws linking the RER to the human mental realm.

And the link laws do not belong to the RER; they are fundamentally outside it. The link laws are contingent; at very least, they are contingent on the existence of some human minds to link *to*. The components of the RER are all essential to it, though, since a reality with a different composition is just a different reality. Or, if the link laws are components of the RER, then human minds are essential to the existence of the RER, which seems to be a stronger form of idealism than the one we are aiming at (and, I trust, would not be a view that a physical realist of any kind would want to endorse).

This forces us to recognise that the functional arrangement necessarily involves something above and beyond the RER. Because the functional arrangement - and thus the physical world - is a matter of what is ultimately presented to human sensory activity (for want of a better expression), it includes the link laws. Thus, it can *never* be identical with the RER, regardless of structural similarity. And since the functional arrangement is constitutively sustained by the link laws as well as the RER, we have arrived at our key target proposition; that the physical world is constitutively sustained by facts about human experience (specifically, facts about the way it is disposed to appear to human experience).

---

[125] Since it would depend on the character of the nomological arrangement of the actual world, and by definition the nomological arrangement is 'not theoretically explicable' - see above.

What remains to be established, then, is that no part of ontologically fundamental reality is physical. This is achieved in a way which will feel familiar from CFI; by reference to physical space. Call the view that some parts of the physical world are ontologically fundamental 'partial realism'; what this really means for our purposes is that some physical phenomena are *identical with* constituents of the RER. In effect, what we are dealing with in partial realism is the equivalent of the Leibnizian attempt to escape the final step of the CFI argument (discussed above); our refutation of it can run along similar lines, though an extra step is needed.

There are two key points to deal with; first, we must show that the dependence of the physical world on link laws means its physical space cannot be identified with the space of the RER, and secondly, we must show that no item can simultaneously occupy two separate spaces (remember that the final step of the CFI argument was to show that no physical-space-occupant could be fundamental unless physical space was).

The key to the first of these points is that the link laws are contingent, so the space they create is contingent. A given RER could have an arrangement which did not create a physical space at all. Our RER is arranged a certain way, a way that creates our physical world, but it *could* be arranged in such a way as to create a very different physical world, or none.

We have seen (in chapter 1) that spatiality is a necessary condition of physicality (though as such it can be construed in either a Leibnizian or Newtonian way). Let U and U' be two different possible universes that share an RER, call it R, but such that in U, R is nomologically organised so as to constitute a physical space P, whereas in U', no such nomological organisation obtains. R may or may not include a space of its own, call it S, but if it does then it must do so in both U and U'; if not, then there can be no S in either U or U'. So, quite obviously, S cannot be identical with P, which by definition exists in U but not U'.

For an object O to be a physical item, it must be spatial; it must be an occupant[126] of a space. In fact, it seems plausible to suggest that it must be an occupant of *physical* space. So, if O is going to be fundamental, then either R has a space, S, and O is an occupant of both P and S, or R has no space, and O is not essentially spatial. We'll begin with the former scenario.

---

[126] To be precise, on a Leibnizian picture, occupancy of a space is reconstrued as *constituency* of a space - playing an ontological role in sustaining the space which, from an ordinary observer's point of view, the given object occupies. I don't think this has an effect on the argument at hand, however.

This is where we are required, if Foster is to be vindicated, to show that nothing can be an occupant of two (overlapping or coextensive) spaces, or at least that nothing can be an occupant of two spaces in the required way. I add the second option because quite obviously, an object *can* exist in two spaces for some kinds of space. As a crude example, if I stand in the doorway of a room, I exist partly in the space of the room and partly outside it. The argument I will offer is that all such examples, however initially compelling, do not involve genuinely distinct spaces, but only arbitrarily-selected regions of a larger space which encompasses both.

More compelling examples are not hard to come by. In *Flatland* (Abbott, 1998), there is a scene where the three-dimensional sphere lowers himself through the two-dimensional space of Flatland, thus occupying both the space of Flatland[127] and the three-dimensional space in which Flatland resides. We could equally imagine there being two Flatlands, close enough together in the third dimension of the universe outside them that the sphere could lower himself through both and thus exist in the spaces of both Flatlands *and* in the three-dimensional space external to them.

Now, obviously, these examples involve only *partial* space-occupancy. The sphere lowering himself through Flatland is never wholly inside Flatland - Flatland cannot contain any of his third dimension. However, there is a *part* of the sphere, a two-dimensional slice, which is wholly inside Flatland and thus putatively occupies two different spaces. This is where we might find a problem for Foster.

We are fortunate to have specified ahead of time a feature of genuine spatiality - which, as what is required for a thing to be physical, is the kind of spatiality we are interested in - which Foster does not explicitly draw on. This is its special significance for the parthood relation, as laid out in the final stage of our refutation of compatibilism in chapter 2. There, we saw that if a thing is genuinely spatial - if it occupies a genuinely spatial field - then all its parts must be genuinely spatial and so must everything of which it is a part.

By this definition, then, the spatial field of Flatland is not a genuinely spatial field, because it allows three-dimensional entities like the sphere to 'dip a toe' into it. It allows parts of some objects to occupy it without the wholes to which the parts belong occupying

---

[127] At least, assuming that in this example 'the space of Flatland' can be treated as numerically distinct from (but wholly contained in) the three-dimensional space in which it resides, rather than simply as a two-dimensional region of the three-dimensional space. We shall return to this point shortly.

it. To be more precise, there is at least one object O which has parts belonging to the spatial field of Flatland and parts not belonging to the spatial field of Flatland.

Still, it might be thought that we can develop an example of an object which stands wholly in two genuinely spatial fields. We can imagine two Flatlands which intersect, perhaps by being two faces of a cube, touching along one edge; the intersection will be the one-dimensional edge. And we can imagine a one-dimensional creature (perhaps transplanted from Lineland) existing in the one-dimensional intersection. This one-dimensional creature will thus occupy the spatial fields of both Flatlands (or at least, we can construe the example in such a way that it occupies both rather than neither).

So far so good, but setting out the example in this way misses a key detail; the two Flatlands overlap, but not exhaustively (that is, there are points in each which are not points in the other). So each Flatland has parts which are in the other and parts which are not; so their fields cannot be genuinely spatial. If we want to construe an object as existing in two different genuine spaces, those spaces must be exactly coextensive; every part of the first space must be in (belong to the spatial field of) the second and vice versa.

This does not necessarily require the spaces to be identical; they could have quite different geometrical structures relating their parts, provided each part of each occupied a point in the structure of the other[128]. Many philosophers might want to rule this out as incoherent, or so far in conflict with our ordinary understanding as to be functionally gibberish, but I cannot see a way to rule out the possibility tout court, so I feel obliged to investigate it.

Physical Space and Fundamental Space

The question comes down to whether P and S can be exactly coextensive but numerically distinct, and whether any object can occupy both as a result. Now, there may be a plausible line of argument against the first of these questions based on taking coextensivity of spaces to involve all regions of each occupying regions of the other and arguing that, since P-regions are ontologically derivative of S-regions, S-regions cannot occupy P-regions, but this would be a strange and complex strategy to develop. The more promising route for us to take is to focus on the latter question. So, assuming that it is *possible* that P and S are coextensive in such a way as to make their fields genuinely spatial, can anything be an occupant of both?

[128] For example, the two spaces might be incongruent counterparts, like a left and a right hand; the points of one might instantiate a mirror-image of the structure of the other.

The answer, I think, is a clear negative. The existence of P is contingent, so the fact that there are P-occupants must also be contingent. On the other hand, the existence of S is contingent only on the existence of the specific RER, R, to which it belongs, which we can stipulate within our example. S and S-occupants exist in any universe where R exists, but P and P-occupants do not.

Let M be an S-occupant and O a P-occupant. O cannot be identical with M because there is some possible world where there are no P-occupants but in which M still exists. Or, to look at it another way, if O is identical with M, it can only be so in virtue of not being physical in worlds where there are no P-occupants. But if there are worlds in which O is not physical, then O cannot be essentially physical, and so at very least we can tell that no fundamental object is essentially physical.

This consequence is logically equivalent to the consequence of the proposition mentioned earlier that there could be no space S - that the RER, R, could be non-spatial, but nomologically organised in such a way as to sustain P. The relevant consequence of this for our purposes is that if O, a P-occupant, is identical with any part of R, it is not essentially spatial. As we saw in chapter 1, spatiality and physicality go hand-in-hand, so anything that is not essentially spatial is not essentially physical and vice versa.

But if this is the case, as just shown, then for anything physical, the physicality of that thing is a contingent matter - specifically, it is contingent on the link laws, the relationship of the RER to the realm of human minds. And this is precisely what we were after. Before reconciling this with our overall argument about the nature of E, though, we would be remiss if we did not check that the problems which thwarted the CFI version of the argument fail to apply here.

There were two objections to the CFI nomological thesis which Foster failed to adequately deal with. One was the transcendental thesis, according to which the correct intrinsic geometry of the physical world was selectable from a God's-eye view of the situation, thus providing a definite, mind-independent answer to the question 'What is the geometrical structure of the physical world?' The other was that the nomological thesis fails to provide an adequate account of the geometrical structure of the physical world in cases where two conflicting structures are equally-good best candidates for that role.

One need only look at the difficulty involved in translating these problems into the AWFU terminological framework to see how much brighter the outlook is for the AWFU argument. The AWFU argument works by showing that the physical world is constitutively sustained by facts about human experience, and thus its correct geometry, irrespective of

any external geometry, is its apparent geometry. Things may appear different from a transcendental viewpoint, but if this is the case it is only so because the transcendental viewpoint is not a view of the physical world at all.

The other problem, of failing to account precisely for what happens when the functional arrangement could have either of two conflicting geometries, stands, but I am inclined to say that in this context Foster's original dismissal of it regains the force that it lost in CFI. In CFI, Foster tried to dismiss the problem on the grounds that it was acceptable to him for the physical world to have an indeterminate geometrical structure, but in CFI he also required that physical space possess its geometrical structure essentially. Assuming that something cannot essentially possess an indeterminate property[129], this means that in the scenario this objection envisages, there is after all no physical space.

But we saw that this ends up making the first pillar of Foster's argument, that the nomological organisation of a physical space is essential to its possession of its geometrical structure, trivial in virtue of the (not completely uncontroversial) way he defines his terms. This opens up a hole in the CFI argument through which an alert realist might slip. On the other hand, I am not sure it troubles the AWFU argument at all. The AWFU argument does not *require* that a physical space essentially possess its geometrical structure[130], but more importantly the dependence of the physical world on the functional arrangement in AWFU is a much 'thicker' matter than in CFI; in CFI, it is simply that there is no other candidate, whereas in AWFU there is a clear explanation of and mechanism for this dependence. It could never be rendered trivial that the physical world is constitutively sustained at least in part by the laws governing how the RER appears to the human mental realm.

Concluding Remarks

How are we to take this result as far as our overall case is concerned? At the start of this chapter, we set up a distinction between reductive realism and reductive idealism as regards the experiential world, E, and its relation to the fundamental world that conditions it, F. Under reductive realism, we said, E derives from (and depends ontologically upon) F

[129] I take this to be uncontroversial because I take the possession of a shape to involve having a determinate boundary. I am quite happy with the idea that some objects - from a simple, macroscopic perspective, clouds would be a good example - do not possess a shape; at the quantum scale I do not see a problem with the suggestion that nothing will turn out to possess a determinate shape at all. Indeed, given the remarks in chapter 1 about spatiality and quantifiability, it should be obvious that the thoroughgoing idealist is committed to the position that there will be no genuine, determinate shapes, however useful approximations to them might be.

[130] Although I think I agree with the Foster of CFI that physical spaces do essentially possess their geometrical structures.

but not in such a way as to give E any ontological dependence on anything mental. Under reductive idealism, E derives from F, but only by means of some other component which is mental (whether or not F is), thus giving E a partial ontological dependence on something mental.

We can translate Foster's concept of 'the physical world', P, as E. In fact, I have done my best to define E exactly as Foster defines P; I substituted 'experiential' for 'physical' because on our definition of 'physical', there are potentially many physical worlds - many worlds which are spatial, quantifiable and objective, including quite possibly F - and I needed some way of picking out the physical world we are interested in for present purposes. As noted above, this has resulted in a slightly unusual use of the term 'experiential', but not, I think, an indefensible one.

Foster's argument in AWFU shows that, assuming the human mental realm is fundamental, P must depend on it. Barring a lengthy and complicated foray into the philosophy of mind - one that by rights should take several books' worth of writing just by itself - we cannot hope to prove that the human mental realm is fundamental. But we are not at quite the impasse this suggests. The requirement that the human mental realm is fundamental is rather stricter than Foster's argument needs *for our purposes*.

This is because we have already (provisionally) accepted that we lack the epistemic resources to argue in any great depth about what *is* fundamental; for present purposes, we are arguing merely about what *cannot* be fundamental, and how those things which are not fundamental are constituted. What we have shown is that facts about the human mental realm M (specifically, about the relation of F to M) are involved in the constitution of E.

All that is required, then, to show that M is ontologically prior to E, is that the ontological dependence demonstrated by Foster's argument is not reciprocal; that there is not also some dependence of M on facts about E. It might be argued, for example, that rather than E in some way depending on M, E and M are ontologically bound up together, neither able to emerge from F without the other - that physical mechanisms which are part of E have some part to play in the constitution of M. The most obvious way of achieving this would be to argue that perceptual mental states are caused by elements of E[131].

This misses a key logical difference between M and E, however. This is roughly that E is contingent while (or in a way that) M is not. Foster's argument relies on the fact that the

---

[131] On Foster's model, perceptual mental states are caused and governed by the link laws, the laws which relate F to M, though the world they *represent* is E.

relationship between E and F is contingent - founded on nomological principles which could be different than they are. Thus any world fitting into the role of E *could be* replaced by some other world, if those laws were changed[132]. So the universe in which the relationship between M and F sustains a different experiential world is entirely logically possible.

Now, it is possible that F could be related to some other mental realm than ours; our existence, after all, is not a necessary state of affairs. As an example, imagine that at some point in the future, a human speciation occurs and an offshoot species, humanas, comes into being. And let us imagine that the major difference between humanas and humans[133] is that the relationship between F and M', the humana mental realm, is governed by a different set of link laws.

Everything about this scenario is conceivable, but it is also not of interest to us. Even if humanas inhabit an experiential world qualitatively identical to E, it is a world for them, not for us. It is not the E we are interested in. This is Foster's last great masterstroke; the dependence of E on the connection between F and M is *what makes it the world whose ontological status we are interested in*[134]. E just is the world constituted in part by the link laws between F and M. So E has an essential ontological dependence on M, and not the other way round.

And so, reductive realism is false and reductive idealism true. Discounting, as we have thus far, naive realism and thus the possibility of identifying E (and M) with F, E ontologically depends (at least partly) on M. From an idealistic point of view, this is a very satisfying result - Foster was content to regard it as a refutation of realism - and given our remarks about the fate of the spatial strategy, it may be as far as we can go. But we should at least take a look at the relationship between E and F, in case there are epistemic resources within it which may allow us to go further. This will be the task of the following chapter.

---

[132] Remember that, by definition, E is the world that appears to us (in Foster's terms, the world as nomologically arranged).

[133] Besides whatever accounts for the fact that interbreeding between the two species can no longer produce fertile offspring, which is the normal biological definition of a speciation.

[134] This is not to be confused with the simpler and more trivial point that what makes E *our* world is its being related in a particular way to human minds. The key issue is that, as we have seen, any physical world related to us in the particular way specified must get its physicality partly in virtue of its relationship to us.

Appendix: Attenuated Causal Structuralism

At the end of chapter 4, we left two theories of causal structuralism open on the grounds that the arguments against them could be better pursued after the conclusion of Foster's case. We are now in a position to return to them. The first of these was a Newtonian 'partial' structuralism, according to which the fundamental ontology of the world is a set of causal power-instances and a set of spatial points at which they are located; the second was a weakened form of the Leibnizian structuralism discussed in chapter 4, according to which space is constituted by spatial rather than causal relations among power-instances.

Given the discussion just completed, the problem with the Newtonian 'partial' structuralism should perhaps be obvious. The assignment of causal powers to spatial points is an intrinsic matter, but it will be the causal powers alone that bring about our perception of the physical world, and there is nothing to guarantee that they should do so in a way which actually represents their intrinsic locations. To put it another way, they may be functionally arranged to appear at very different locations to where they intrinsically are.

The structuralist may choose to argue that a match between functional and intrinsic arrangements is a better explanation than any case of deviance, but this misses the point. The key point is that a good match is a better explanation *only in virtue of the fact* that the functional arrangement is how things appear to us - and whether or not we ourselves can be reduced to collections of causal powers, this is still a fact about something mental.

The problem for the attenuated Leibnizian view is similar. Here, space is sustained by a set of non-causal relations among power-instances, but the power-instances get their identities from their causal relations - their places in the causal structure. Putting aside the question of whether this requires (I think rather against intuition) that space has an asymmetrical structure, we can see that again, the space that is presented to us in experience must be a matter governed at least in part by the causal structure. So again it is possible that our experiences could be nomologically organised to be deviant from the structure fixed by the intrinsically spatial relationships.

6. Naive Realism and Relativity

Scientific Resources

In the last chapter, we ignored naive realism to focus on the debate between reductive realism and reductive idealism. This step was taken because the business of arguing against naive realism on purely philosophical/logical grounds is tricky, particularly given the blunt stubbornness of some significant naive realist arguments (as, for example, those of G.E. Moore). However, as we move from considering E in itself to considering the relationship between E and F, we may find new lines of argument opening up to us.

I shall begin with the assumption that among the key interests of most if not all forms of realism is a desire to take seriously the utility of physical science (or perhaps to take seriously the possibility of a complete and correct final theory of physical science). I happen to believe this desire is far from incompatible with plausible forms of idealism, but I recognise this will go against the intuitions of many realists.

While naive realism is most often discussed as an epistemic theory, according to which (roughly) secondary qualities inhere in objects *as they are perceived*, rather than as powers to produce certain appearances under certain conditions, the naive realism I am addressing here is a more strictly metaphysical theory. It is the theory that the world as it appears – the experiential world – is fundamental, and that the postulates of physical science are functional constructs only.

In short, it is the view that E and F are numerically identical or at least not ontologically separate, as discussed in the previous chapter. We saw there that Foster's anti-realist argument turns on the nature of the process by which E is constituted by F, and thus that it *assumes* that E is constituted by F. A realist, therefore, may hope to preserve his position against Foster's attack by denying this constitution, and saying that E and F are ontologically indistinguishable. This, combined with the assumption that realists wish to uphold the utility of physical science, constitutes the position I will argue against in this chapter.

There are many much-discussed arguments against naive realism; whatever its attractions as an account of perception, it seems to have a number of uncomfortable metaphysical consequences. The argument I shall present here is simply the strongest I have been able to come up with. It turns on a matter of what physical science, taken seriously, says about the relationship between E and F. I take the naive realist to be

committed to the assertion that E, the experiential world, the world of common-sense macroscopic objects, just is F, the ontologically fundamental world.

Crucially for our purposes, naive realism is committed to taking anything which can feature in a human experience as being a possible constituent of the fundamental world. I will argue that physical science, and specifically the special theory of relativity, shows that if this is true, then it must be possible for contradictory states of affairs to obtain. In other words, in attempting to do justice to physical science, the naive realist will find himself inescapably committed to a contradiction.

Relativistic Time Dilation

Two Basic Premises

The special theory of relativity derives from only two fundamental premises, and can be explained very neatly in terms of them with the benefit of only a very little mathematics. Since the logic of it is deeply elegant, it is worth going through in some detail.

The first of the two premises, from which the theory gets its name, is *relativity*. More precisely, it is the Galilean hypothesis that there is no absolute frame of reference for motion; that the motion of an object can only be defined *relative to* some other object. When we consider objects in ordinary life to be at rest or unmoving, it is only because we judge them to be moving at the same speed as the majority of the rest of the environment (by way of a simple illustration, the house in which I write this is at rest relative to the part of the surface of the Earth on which it stands; but relative to the Sun, both house and planet are travelling at some sixty-seven thousand miles per hour).

What we need, though, is a proposition a little stronger and narrower than the Galilean hypothesis. This is the proposition that when an object is not accelerating (i.e. when it is travelling at a constant speed in a straight line), there is absolutely no difference at all between saying that the object is moving and saying that the universe is moving past the object. There is literally *no fact of the matter* to make one or the other interpretation preferable.

This is, of course, at once quite familiar on a case-by-case basis and quite counterintuitive when taken as a universal truth. However much we understand that, counter to appearance, the Earth goes round the Sun, and spins on its axis, we still ordinarily think of some things as being 'at rest'. The concept of stillness is a fundamental part of our normal conceptual scheme, because it is very rare for us to have to think in terms of things which do not (at least as far as the relativistic scale is concerned) move at

roughly the same speed as us. The effects of the special theory of relativity manifest detectably only when relative velocities reach noticeable fractions of the speed of light.

Which brings us neatly to the second premise of the theory. This premise derives from Maxwell's equations describing electricity, magnetism, and the electromagnetic field. The equations are complicated and we don't need to know the details of them for our purposes, so I won't state them here. All we need to know is that the equations are extremely accurate, covering ordinary phenomena perfectly, and that they feature a constant, c, which is the speed at which electromagnetic waves travel in a vacuum.

The problem is that the speed of electromagnetic waves (or, as it's more commonly known, the speed of light, approximately 300,000,000 metres per second) is defined in Maxwell's equations as an absolute value. There is no room in the equations to say what the speed is relative to. But this directly contradicts our first premise, that all motion is relative to something.

Enter Einstein. His brilliant step forward with this problem was to investigate what would happen if he took electromagnetic waves to travel at the same speed relative to *all* other objects, regardless of their motion. If you are moving in the same direction as a beam of light, however fast you are going, the front of the beam of light moves away from you at roughly 300,000,000 metres per second; if you turn around and run the other way, however fast you are going, the beam of light moves away from your back at roughly 300,000,000 metres per second. And despite the fact that your velocity has changed, perhaps drastically, the beam of light does not decelerate in order for this to be true.

This seems like gibberish, but stranger results are yet to come. We don't notice these effects because the speeds we operate at in ordinary life are so small when compared to the speed of light. We walk at about three or four miles an hour, which is about a metre and a half per second. Cars on a motorway, at 70mph, are doing just over 30 metres per second, or one ten-millionth of the speed of light. Concorde's best cruising speed was about 600 metres per second, two millionths of the speed of light[135]. As such, scenarios in which we might *notice* light behaving in a way not compatible with our ordinary

[135] Clocks exist which are sensitive enough to detect relativistic distortions at this speed; analogues of some of the examples I am about to discuss have been conducted using supersonic jets and clocks of this type.

expectations of how velocities add up and subtract are limited to the purely hypothetical[136].

The Light Clock

Everything we need to know about relativistic time dilation can be explained in terms of two thought experiments and the premises just given. The first of these is the 'light clock'[137]. The light clock consists of a pair of mirrors, a metre apart, with a beam of light bouncing back and forth between them. The clock is considered to 'tick' every time the beam of light returns to one mirror - that is, each time it completes a round trip of two metres, from one mirror to the other and back again.

The speed of light is very nearly 300,000,000 metres per second[138], so the light clock ticks 150,000,000 times every second. One tick takes roughly 6.67 nanoseconds (0.00000000667 seconds). This is quite important, because, given our first premise above, we know that this figure is fixed regardless of what speed the clock is travelling at; if it was not, then if we had a light clock in the room with us, we could tell just by measuring its ticks whether we were in motion or at rest.

So far, so unproblematic. We naturally assume that a clock ticks at the same rate regardless of whether it is moving or stationary. But the advantage of a light clock is that we can study in detail what happens if we put it in motion - for instance, by mounting it on a fast train (travelling at about 120kph).

Now, if you've ever tossed a ball vertically upwards on a moving train, you'll know that the ball comes back to your hand as if you aren't moving - a prime demonstration, by the way, of Galilean relativity. The ball behaves as if it travels vertically upward, and vertically back down, despite the fact that from the perspective of somebody standing beside the tracks, outside the train, you may have moved many metres in the time, turning the ball's motion into an arc.

Something similar goes for anything else that moves in a 'straight line' while on a moving train. Sit on the opposite side of a carriage from a friend and toss a ball back and

---

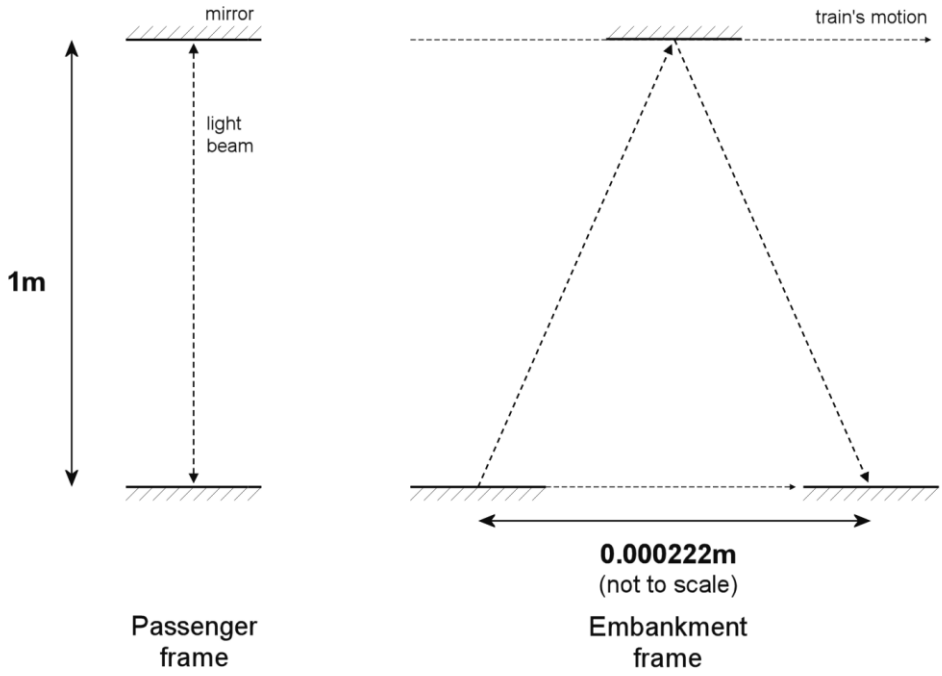
[136] This is not to say that there aren't technological devices which need to take account of these effects, such as satellites in low Earth orbit, just that any such device must be more sensitive than human perception, as of course many are.

[137] This is a long-established introductory thought experiment for relativity. The particular treatment of it I am giving is largely due to Cox and Forshaw (2009).

[138] I use the approximation rather than the precise value, 299,792,458m/s, purely to simplify the maths a little.

forth between you; you need make no compensation for the train's motion[139]. So it is with the light clock; from the perspective of someone on the train, the beam of light travels one metre to the second mirror, then back along the same path one metre to the mirror it started from. The total distance travelled is two metres, so it is no surprise that the single tick takes 6.67 nanoseconds.

But from the perspective of an observer standing on the embankment as the train whistles past at 120kph, the picture is ever-so-slightly different. Just as the path of the thrown ball becomes an arc for this observer, the light beam turns out to travel on a diagonal. The angle of the diagonal is very steep, because even at 120kph, 6.67 nanoseconds of travel doesn't take you very far. In fact, it takes you only 0.000222mm. But this still means that the beam of light travels slightly *further* than two metres.



A single tick of the light clock, modelled as if from the perspective of a passenger (on the left) and from the perspective of an observer outside the train (on the right).

Pythagoras' theorem gives us the difference. In what is called the train's *frame of reference* - that is, how things appear to be from the perspective of an observer on the

[139] Not to be confused with its acceleration; if the train is accelerating or decelerating, you'll need to compensate for that, but what we're interested in is steady motion, not changing motion.

train - we know that the beam travels two metres. In the frame of reference belonging to the observer outside the train, the beam travels a distance equal to $\sqrt{(2000^2 + 0.000222^2)}$mm, which is $\sqrt{4000000.000000049423}$mm, a value just slightly larger than the two metres the beam travelled in the train's frame of reference.

The difference is almost insignificant. But it is significant for our purposes. We have started from the standpoint of a constant speed of light *and* a constant tick duration between frames of reference - that is, we know that the beam of light must travel just as fast to make its journey, and take just as long, whether we view matters from inside or outside the train. But this is impossible, since from the perspective of the observer outside the train, the distance the beam travels (call this the 'tick distance') is greater.

Imagine that we speed up the train so that we get a significant difference between the tick distance on the train and off it. If the train travels at half the speed of light, for example, it will travel a full metre in the time it takes the light clock to tick. So the tick distance for the frame of reference for the observer outside the train will be $\sqrt{(2^2 + 1^2)}$, which is $\sqrt{5}$, roughly 2.24m. We know how long the tick takes in terms of time - 6.67 nanoseconds - so we can easily work out the speed the beam must have travelled at.

It turns out to be travelling at about 335 million metres per second - a significant increase over the speed of light, which we were taking as constant. How do we resolve this? By allowing the seconds themselves to stretch. To put it another way, time on the train runs slower than it does on the platform[140]. This is the basic form of relativistic time dilation.

Order Flexibility

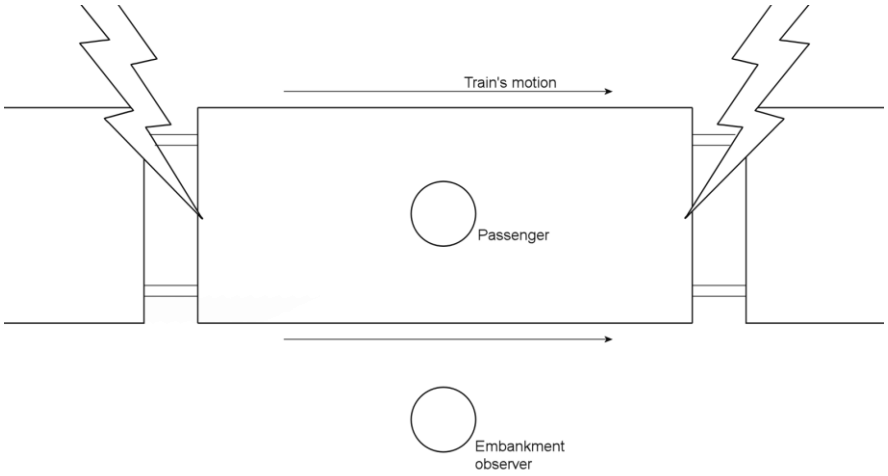
So far, we have not generated a problem that the naive realist cannot handle with only minor adjustments to his position. Technically, the experiences of the observer on the train and those of the observer outside do conflict - one experiences more time than the other - but they count the same number of ticks and so on, so there is very little concrete within their experiences to draw out a contradiction.

However, there is another, more drastic feature of special relativity which does generate flat-out contradictions between the experiences of different observers. We will

[140] This leads to the famous 'twin puzzle', where one of a pair of twins becomes an astronaut, makes a long journey at a noticeable fraction of the speed of light, and when he returns to his twin, the twin has aged much more in the intervening time. Lest it be thought that the effect is purely hypothetical, experiments of this nature using hypersensitive clocks and supersonic jets have been performed, and have confirmed that time dilation occurs.

discuss in due time whether this is a truly fatal blow to naive realism (I am of the opinion it is), but for now we should outline the scenario. The phenomenon we are interested in is a set of scenarios in which the apparent temporal order of events changes depending on the velocity of an observer relative to them.

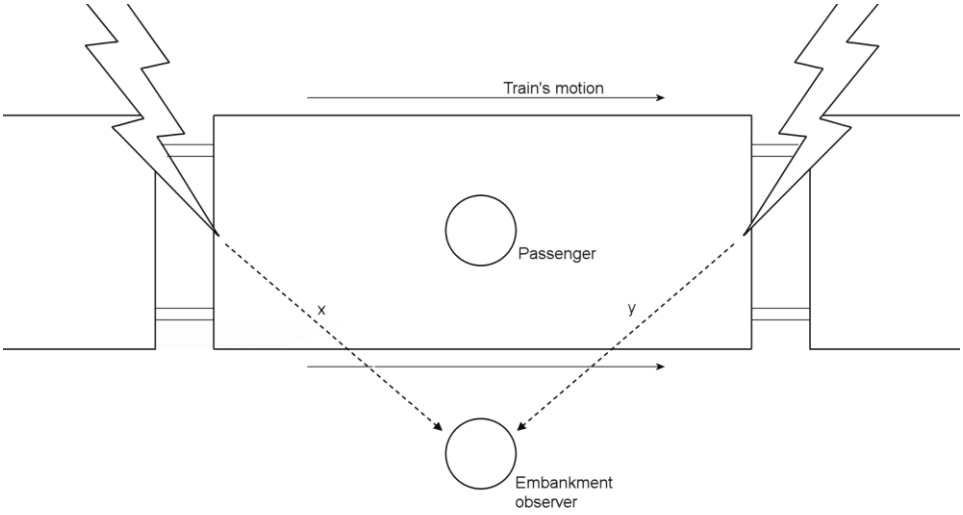
In this, we can follow a thought experiment taken directly from Einstein (1916). Sticking with our example of the train, we can call the frame of reference belonging to the observer on the train the *passenger frame*, and the frame belonging to the observer outside the train the *embankment frame*. Now imagine that, from the perspective of the embankment frame, at the instant at which the two observers 'line up', a bolt of lightning strikes each end of the train in such a way that the strikes are visible both inside and outside the train.



Lightning strikes both ends of the train carriage at the moment when the passenger and embankment observer are 'level' - that is, both are equidistant from the ends of the carriage.

What happens from the perspective of the embankment frame is that light starts to propagate from the strike sites at the instant of 'lining up', and reaches the embankment observer shortly afterwards. Knowing the speed of light and the distance to each end of the train[141], the embankment observer can extrapolate backwards from the point at which the light reaches him and work out that the two strikes were simultaneous - each beam of light travelled an equal distance at an equal speed, and so if they arrive simultaneously the events must have been simultaneous.

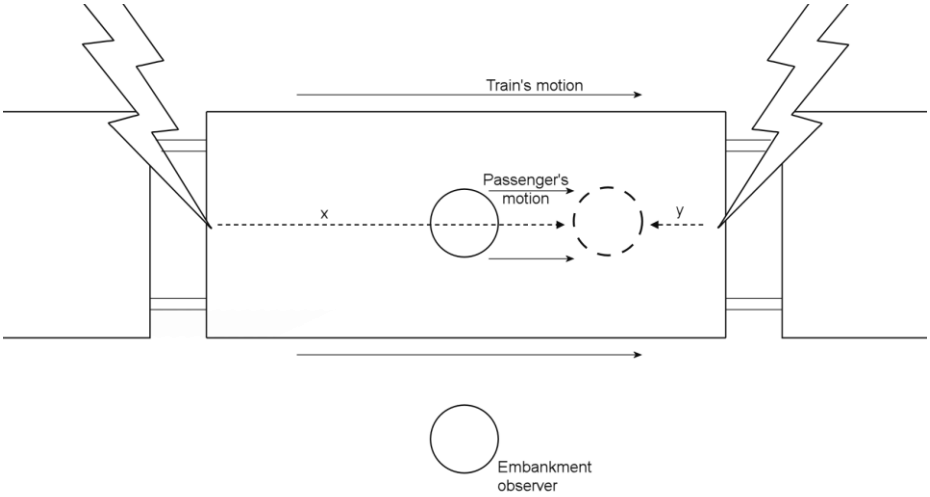
[141] For simplicity's sake, we can stipulate that the embankment observer is equidistant between the two lightning strikes; the key point is simply that he can extrapolate that the two strikes are simultaneous.



Assuming that the embankment observer somehow knows he is equidistant between the two lightning strikes, he knows that the beams of light x and y must be the same length. Since the speed of light is a constant, he therefore knows that the timing of the strikes must match the timing of the arrival of the light - that is, the strikes must be simultaneous.

So far, so unproblematic. But now consider matters from the perspective of the passenger frame. Remember that the passenger observer has no way of telling whether he is in motion (and the embankment observer at rest) or at rest (and the embankment observer in motion, as the world rushes past around the train). The lightning strikes both ends of the carriage, and light starts to propagate along the carriage towards the passenger.

But the passenger *is* in motion relative to the lightning strikes, in that he is moving towards one and away from the other. He has no way of knowing this, because the strikes, we can stipulate, are momentary events, but the upshot of it is that the light from the strike ahead of him has *less far to go*. So it arrives sooner than the light from the strike behind.



> *Because the passenger is moving towards one of the strikes, the distances x and y differ, so the light from the strikes reaches the passenger at different times. However,* to the passenger*, x and y* appear *to be the same length, leading to the judgement that the strikes themselves were not simultaneous.*

The precise maths are beyond our concern, and in fact involve rather more complicated effects than those I am about to describe (for example, neither of the lightning strikes is actually simultaneous with the moment the passenger and embankment observers 'line up' in the passenger frame). I will keep it simple, and focus on the process the passenger must go through to work out *when* the lightning strikes happened.

He starts with the datum that he saw lightning strike the end of the train ahead of him, then a moment later the end of the train behind him. He does not know whether the train is in motion at all, so he cannot use the information - accessible to the embankment observer - that the train has a certain effect on the distances to the points of the lightning strikes. All he knows is the distance the light must have travelled to reach him from each end of the train[142]; and these distances are equal.

Taking the speed of light as constant, then, he knows that the strike at the front[143] of the train must have happened first - how else could light from it have reached him sooner across an equal distance than the light from the rear strike? In the passenger frame, the two lightning strikes are *not simultaneous*.

---

[142] Again, we'll assume he's in the middle of the train for simplicity's sake.

[143] We have to be careful here, since technically the passenger cannot tell that the train is in motion, so 'front' might be considered a contentious notion. For our purposes, we could just label one strike x and the other y, such that x is the strike which appears to be ahead of the passenger in his direction of motion according to the embankment frame, and which appears to the passenger to happen first. 'Front' here is a linguistic convenience and nothing more.

It is very hard indeed to shrug off the intuition that this is just because the passenger frame is in motion and the breach of simultaneity can be explained away by that fact. The crucial point about special relativity is that this is *not* the case. The world furnishes us with *no* epistemic resources whatsoever to decide that one of the frames should be taken more seriously than the other.

If further reinforcement is needed, consider briefly these two variations on the scenario: First, imagine that at the same time that the passenger and embankment observer line up (according to the embankment frame), a second train is travelling in the opposite direction, with a passenger of its own sat in the middle of a carriage, who also lines up (according to the embankment frame) with the first passenger at the same moment. It should be clear from the foregoing that to the second passenger, the lightning strike which is at the back of the first train (relative to its motion in the embankment frame) will occur *before* the strike at the front[144].

Now strip away the planet and the embankment frame altogether. Replace the trains with spaceships, their passengers with astronauts, and the lightning strikes (which after all could spell very bad news for our astronauts) with astronomically distant events - supernovae, perhaps. Call the supernovae A and B. Now, we have just two frames, each moving relative to the other and to the supernovae. According to one frame, A goes supernova before B; according to the other frame, B goes supernova before A.

This last example is useful not only because it is in some ways more plausible, but also because it strips away the illusion of a 'preferred frame' of being stationary relative to some point on the Earth's surface. It shows very clearly that the temporal order of things depends on the speed one is going; more importantly for our purposes, it shows that two conflicting propositions are true[145]:

P: A's going supernova appears to precede B's going supernova
Q: B's going supernova appears to precede A's going supernova

Thus formulated, P and Q are quite obviously contradictory. The special theory solves this problem by saying that temporal order reduces to a deeper phenomenon, intervals in four-dimensional spacetime, by a process some parts of which are quite well

---

[144] I apologise for the density of qualifiers needed to explain relative terms like 'front' and 'back' in relativity theory; they are, unfortunately, a necessity.

[145] In the example, at least. For our purposes it is enough that such examples are physically (and not merely logically) possible.

understood[146]. Our question, then, is whether the naive realist can avail himself of any such mechanism.

Contexts in Naive Realism

We ordinarily take it that a theory which entails a contradiction must be false. Naive realism appears to entail the contradiction of P and Q. So the question is whether the apparent contradiction is as it appears to be (I will refrain from the perhaps trite argument that the naive realist is committed to things being as they appear to be - we can allow him at least some form of strategy to deal with non-veridical perception).

The obvious strategy for the naive realist is going to be to appeal to some sort of context operator; that is, to say that P and Q are not accurate statements about what occurs in the above examples because they do not account for the fact that temporal order is relative. This strategy replaces P and Q with:

P': A's going supernova precedes B's going supernova in x's frame.

Q': B's going supernova precedes A's going supernova in y's frame.

Where 'x' and 'y' pick out the observers aboard the passing spaceships. So is this approach viable?

The simplest way to take this approach is to interpret frames of reference as just another feature of points of view. Temporal order-switching as outlined above then becomes no different to the way that, for example, Radio City Tower may appear to be to the left of the Metropolitan Cathedral tower if one looks at it from South Liverpool, but will shift to the right of it as one moves northward. We need not be so brutish as to deny that naive realism can handle visual field place-swapping of this kind.

But visual field place-swapping has a feature that at least some versions of order flexibility do not (at least under naive realism). Specifically, there is a way to describe the relative locations of the Cathedral and the Radio City Tower which makes no explicit reference to any observer (or even the potential point of view of any observer). We can say that the two landmarks are so many metres apart as the crow flies, and specify which

[146] In the sense that we can predict how a four-dimensional spacetime interval will translate into its spatial and temporal analogues for any hypothetical frame of reference. The precise scientific details of the system underlying this translation are, as I understand it, still a subject of debate, but the formulas that describe it are extremely precise.

direction the crow would need to fly that many metres in to reach one from the other. We could also give their coordinates in any legitimate latitude-longitude base. Given the 'no-nonsense' metaphysical aspirations that tend to go hand-in-hand with naive realism, we can regard this 'objective' description as being more accurate, perhaps even more correct.

Attempt to do this in the case of order switching, and we just jump back from P' and Q' to P and Q. The temporal interval between the two events differs between observer frames[147]; eliminate the frame-relative language and we are left with only the contradiction, in what is supposed to be the 'more accurate' description. This is far from refuting naive realism, but it does show that the naive realist cannot rely on this simple version of the context strategy.

Naive Scientific Realism

Instead, the only serious option I can see for the naive realist is to accept the scientific explanation of order switching. This version of the context strategy puts rather more weight on the back of the context operator. The naive scientific realist says, in effect, that P' and Q' should be replaced again, this time with statements about non-veridical appearances. To whit:

> *P'': X mistakenly perceives[148] A's going supernova as preceding B's going supernova.*
>
> *Q'': Y mistakenly perceives B's going supernova as preceding A's going supernova.*

This takes some explaining. The special theory explains order switching by saying that all apparent temporal (and spatial) intervals between two events are differing manifestations of a fixed spatiotemporal interval, and the speed of a given observer

---

[147] Or, if in the case of the two spaceships moving in opposite directions, you are not happy that this is necessarily the case - and certainly there must be some pair of relative velocities of the ships according to which both extrapolate the same interval between the two supernovae, but with the order running in opposite directions - we can go back to the earlier train example and consider the picture where two events are simultaneous in one frame and not simultaneous in another. Here the temporal intervals perceived in the different frames are ineliminably different between frames; one interval (between the same pair of events) cannot simultaneously be simultaneity and non-simultaneity.

[148] According to some naive realist terminological frameworks, it might be a mistake to use the word 'perceives' here (in that they would prefer to think of non-veridical perception as not actually being perception), but we can adjust P'' and Q'' to any preferred terminological framework, I think, without affecting the points I am about to make.

relative to the events determines how much of the spatiotemporal interval manifests as time rather than space for that observer.

But the naive realist says that all instances of veridical perception are instances of direct contact with the fundamental world (or at least, the version of naive realism we are interested in - of which more later). So if in the fundamental world events are separated by spatiotemporal intervals, then any perception of a purely temporal interval[149] must be non-veridical.

The absurd (or at least unpalatable) consequence of this is clear; if he chooses to pursue the scientific strategy, the naive realist makes *all* perception of temporal intervals non-veridical (because every temporal interval is derived from a spatiotemporal interval). In fact, he makes all perception of spatial intervals non-veridical too, because relativity theory says exactly the same things about space as it does about time (i.e. that it derives from an underlying union of the two).

Exactly how widespread and damaging one considers this to be depends on fine details of how one understands naive realism. Specifically, it depends on how much of our ordinary perception one considers to involve (either be of or entail) spatial and temporal intervals, which will depend on what one takes to be the fundamental objects of perception (atoms, parts of objects, macroscopic wholes etc.). This will vary from model to model, but as a minimum, all spatial and temporal beliefs derived from perception will turn out to be false.

It may still be possible to build a consistent system of naive scientific realism. After all, relativity theory provides a complete and consistent system describing how the non-veridical temporal and spatial appearances derive from the underlying spatiotemporal relationships. I am not sure any actual naive realist would welcome such a strange position, but we can assume there is a *possible* naive realist who does. Fortunately for our purposes, this view is irrelevant.

---

[149] Do we perceive temporal intervals or infer them? The naive realist might well want to argue for the latter. This will avail him nothing, however, because as I understand it, under naive realism, anything that can be correctly inferred from experience is true of the external world, and it will not be the inference that is incorrect in this case. If x perceives A's going supernova followed by B's going supernova, it is entirely correct for him to infer that A's going supernova preceded B's; the error (distortion of how things stand in objective reality) is in the appearance of A's supernova 'first'.

This is because we are only interested in a certain very narrowly-defined kind of naive realism, one whose boundaries we have already left. We are interested in the form of naive realism which escapes Foster's argument - the only position according to which E cannot be shown to be ontologically dependent on its relationship to the human mental realm. Naive scientific realism, I think, clearly falls within the scope of Foster's argument.

Foster's argument, as I unpacked it in the previous chapter, applies to any world in which E, the world we experience and find ourselves in, is not identical with F, the fundamental world underlying it. The naive scientific realist says that a whole swathe of our everyday experiences are actually non-veridical, and are to be explained in terms of a complex causal process whose underlying phenomena (4-dimensional spacetime intervals) they do not accurately represent.

To escape commitment to a numerical separation between E and F, the naive realist will presumably assert that the non-veridical temporal and spatial experiences brought about by special relativity are not strictly part of E (perhaps by trying to construe them as elements belonging purely to the human mental realm[150]). But this misses the point of the way I have defined E. E *just is* the world that appears to us. If it appears to us that there are spatial intervals and, separately, temporal intervals, then those intervals are features of E, even if ultimately construed as reducing to some (non-experiential) mental items and their non-mental causes.

The naive scientific realist is thus committed to a numerical separation between E and F. But this means that there must be some relationship of constitution between the two (that is, E must be constituted by the behaviour of elements of F), and we saw in the previous chapter that whenever this is the case, there must be a mental element to the process - specifically, the relationship to the mental that guarantees that E is the world we are interested in. Is there any option remaining to the naive realist which threads between

---

[150] Which brings up the need to clarify the distinction between E and the human mental realm a bit more, since it is *not* my intention that what we might in common language describe as 'inner thoughts' feature in E. For example, I consider my thought that this chair is uncomfortable (or that the discomfort of the chair is useful for exemplary purposes) to be entirely constrained to the human mental realm; the cause of the discomfort itself, however - the hardness of the seat - is a part of E. This narrows the sense of 'experience' according to which E is 'the experiential world' since arguably at least some propositional thoughts are *inward experiences*. If we are to adopt such a schema, then I assume we can be relatively clear on an intuitive notion of the difference between 'inner' and 'outer' experiences.

scientific realism, which falls to Foster's arguments, and the order-switching paradox? Perhaps.

To do this - to be what we might call a strict naive realist - one would have to invoke a context operator with no scientific content, one which does not require going beyond experience to explain experience. The only way I can see of doing this would be to stick closely to the naive realist premise that what appears is what is, and bite the bullet of saying that if what appears differs between frames, then different frames experience numerically different things.

Fundamentally, then we are returning to P' and Q', above. The difference this time is that we are abandoning the idea that there is one objectively correct order for the two supernovae - indeed, that there is an objectively correct temporal framework at all. So we are already into territory a naive realist is likely to find uncomfortable. But, prima facie, this strategy saves the appearances. What remains is the question of cost.

And this is where, for our purposes, the earlier light clock example becomes relevant. I can see no reason why the naive realist could systematically treat order switching differently to time dilation; both are, after all, ways in which a temporal framework can differ between reference frames. So we can expect that the strict naive realist will agree that the passenger frame experiences (or that the experiences which constitute it entail) a different length of second to the embankment frame *just because* in the passenger frame, seconds are longer than in the embankment frame.

So far so good, but only because we have not stressed a key point about time dilation. This is that there will be time dilation between any two frames of reference moving at different velocities. The dilation may be miniscule and below the threshold of conscious perception, but it is there. And this puts the strict naive realist in a quandary.

He can hold tight to naive realism, insist that if time dilation has no perceptible effect, it does not occur, and abandon a systematic description of time dilation where it does have a perceptible effect[151], or he can allow that imperceptible time dilation does occur, and suffer the consequence of accepting that a second has different lengths for any two distinct frames.

The former option, I think, is clearly unacceptable. It involves a gross failure to do justice to one of the greatest scientific triumphs of the modern age. Worse yet, it

---

[151] Because the 'correct' (or at least our current best) description of time dilation directly contradicts the assertion that time dilation does not occur unless it is above the threshold of conscious perception.

invalidates the simple, elegant process of extrapolation from experience that we pursued to outline time dilation earlier. Each step of that process was, I think, deductively correct. So if we allow the former option, we throw out the system of deduction which validated our description of time dilation (at least, we remove its completeness). This bullet is rather too big to bite.

What of the second option? The question here is a little less clear-cut, but I still think the consequences are deeply unwelcome. Let us take a hypothetical God's-eye view of the world, and consider how frequently different individuals stand in distinct frames.

We know that any time there is relative motion, there is a distinction between frames. So obviously, if I am standing still, you are walking, and someone else drives past us both in a car, there are three different frames involved. What we are really interested in is what happens if you and I both stand 'still'; is it ever feasible to describe us as at rest relative to one another[152]?

It might be thought that the answer is 'yes, if we stand close enough together'. After all, if we stand on opposite sides of the rotating Earth, we will clearly be moving in opposite directions all the time. On the other hand, if we stand in the same (sensibly small) room, we ought to find ourselves to be moving at the same speed.

This is not the case, though. It seems that way because the differences in speed at distances of a few feet become ridiculously small, but the truth is that almost any two distinct points on the Earth's surface will always be moving at different velocities. Consider first what happens if you stand to the South of me while we are both North of the Equator.

In this case, you will be closer to the Equator than me. The Earth's surface, quite naturally, is much further from its rotational axis at the Equator than anywhere nearer the poles. So, in the course of one full rotation of the Earth, you will travel further than me, but in the same amount of time (assuming the rotational speed of the Earth is precisely fixed). So you will be travelling (albeit fractionally) faster than me.

It is harder to explain what happens if you stand due East or West of me, but in essence we turn out on close analysis to be orbiting one another. Remove the Earth from the picture and just consider what would happen to us as we transcribed an Earth-sized circular path through space over the course of a day. Twelve hours after we started out, we would be moving in the opposite direction to the direction we started out in, but you

[152] That is, moving at the same velocity - the same speed in the same direction.

would still be ahead of (or behind) me. Another twelve hours later, we would be back to roughly where we started, and you would still be ahead of (or behind) me.

Combining East-West and North-South differences in position will do nothing to stabilise our common frame. There may be points on the Earth's surface where it is possible for two people to stand in such a way that they have no relative motion[153], and there certainly are scenarios in which we can envisage two people having no relative motion, but these cases will be the exception, not the rule.

The harsh truth for the strict naive realist is that he will be forced to acknowledge that in ordinary life, people constantly exist in distinct frames, and thus have seconds of distinct length - and, it is worth noting, centimetres and inches of distinct lengths, because relativity theory also produces spatial dilation between differing frames. This is strange enough, but there is a deeper problem at work.

We seem to be able to communicate about the spatial and temporal intervals between things. When I say 'Usain Bolt's record for the 100m sprint is 9.58 seconds', you know what each of those metres and seconds amounts to. But on strict naive realism, we are never acquainted with the same metre and second at any given point in time[154]. So what explains the possibility of communication?

Clearly, some common frame of reference is required. But the common frame of reference *is just* how things stand in four-dimensional spacetime and the strict naive realist cannot avail himself of this set of imperceptible facts. What is true is what appears, but what appears contradicts the existence of an underlying four-dimensional spacetime. To invoke spacetime would be to go back to naive scientific realism, but we have seen that that runs into Foster's argument. The naive realist can either have a systematic theory of relativity or ordinary language about time and space, and never both.

So there is no way to account for all features of E without appeal to something standing behind it. And if one appeals to something standing behind it - F - one is forced by Foster's argument to acknowledge that E depends on its relation to human minds. So, whatever the status of the fundamental world, some form of idealism is true of E, the world we most naturally consider ourselves to inhabit.

---

[153] For example, if you are closer to the equator than me, but I am uphill from you on a slope which precisely compensates for the reduction in my distance from the Earth's rotational axis, we may find that we travel exactly the same distance in a day. As a simpler point, we might both become intangible and occupy the *same* point. But these are clearly both circumstances far outside the norm.

[154] Strictly, in spacetime, though it's not clear that the strict naive realist can describe it thus.

7. Sketch of the Next Step

Showing that we should be idealists - specifically, reductive idealists - about E is a satisfying point at which to bring the present project to a close. It is therefore time to review the arguments that brought us to this point and examine what might take us from our present position to a full-blooded idealism about the fundamental world.

We began by laying out a plausible definition of the terms 'mental' and 'physical'. By this definition, the 'mental' is that which is subjective, qualitative, and non-spatial; the physical is that which is objective (or at least non-subjective), quantitative and spatial. We defined these terms in a plausible (though perhaps not completely standard) way and showed that each triad of properties was coextensive.

This enabled us to show that any given object must be either mental or physical - there is no middle ground - and that nothing can be both. With the aid of a more tentative extension to our definition of spatiality, to do with parts of things and their membership of genuine spatial fields, we showed that nothing can have both physical and mental parts either.

These three points combined into a refutation of a set of positions I termed *compatibilism* - attempts to treat the mental-physical divide as in some way inconsequential or unimportant. This left us with three positions: (substance) dualism, (physical) realism and idealism. We briefly considered and rejected two versions of the spatial strategy for idealism, Kant's and Bradley's, and showed that there is a significant problem of epistemic access to fundamental reality, which bars most arguments analogous to those of Kant and Bradley.

From this unfortunate result, we moved into a discussion of the experiential world, E; the world we prima facie appear to inhabit, the world of common-sense macroscopic objects such as tables and chairs. We laid out three possible attitudes we could take to E: naive realism, according to which E itself is fundamental, reductive realism, according to which E is non-fundamental, but constituted in a way which does not involve any minds or mental facts, and reductive idealism, according to which the constitution of E necessarily introduces an ontological dependence on some mental things.

We discussed John Foster's arguments and showed how they could be used to establish that, if we ignore naive realism, reductive idealism is our only option. That is to say, we proved that if E is not fundamental, it must ontologically depend on some mental facts. Specifically, we proved that any given world can only be E (for us) if it gets its

physicality in part from the relationship between the reality underlying it and human minds.

Finally, we have shown that the forms of naive realism which are strict enough to escape Foster's arguments about E are all either self-refuting or so absurd as to be practically untenable. We did this by means of two puzzles about time arising from the special theory of relativity; time dilation and order-switching between reference frames. This has led us to the conclusion that reductive idealism is true of E.

However, there is a gulf between reductive idealism about E and what should truly be called idealism, namely the view that *fundamental* reality must be wholly mental (or wholly non-physical, which, given the definitions established in the first chapter, is the same thing). The difficulty that arguments about fundamental reality face is, as we saw in our discussion of the spatial strategy, that our epistemic resources concerning fundamental reality are so scarce as to make any argument about it dubious[155]. We could never be sure that some fact we *cannot* know about fundamental reality stands in conflict with any of the assumptions our argument might require.

However, all is not necessarily lost. There is an argument, which we do not have space to pursue in detail here, which may escape this problem, at least if we allow (and we perhaps should not) that at least our formal logical framework may be applied to the fundamental reality. This is because it is an argument based on components of E and what is or is not suited to constitute them.

In our discussion of Foster's arguments, we showed that causal structuralism is false - that E cannot be exhaustively described purely in terms of causal powers and dispositions. This means that E, or parts of it at least, must have some intrinsic content properties. And this forms a basis for our line of argument (in fact, it may be the most difficult step, already conquered). The argument goes like this:

1. E has some intrinsic content properties.
2. Intrinsic content properties cannot be constituted by a purely structural fundament (to put it another way, the fundamental reality must also have some intrinsic content).

---

[155] This may encourage some to reject the existence of the fundamental, underlying world altogether on parsimony grounds, and develop a strictly phenomenalistic theory which describes E as a systematic pattern of experiences. I see nothing wrong with such a theory, but I do not think the parsimony argument is strong enough alone to support it, so I have not discussed the possibility in depth.

3. To be an intrinsic content property, an item must have certain properties over and above its intrinsic character, such as being causally inert, which in combination will entail that it will fit our definition of 'mental' rather than 'physical'.

This gives us the result that the fundamental world must be intrinsically mental, which should be enough to satisfy all but the most extreme idealists. I take us to already have proved premise 1, or at least to have proved something so close to it that what remains is only a small step. Premise 2, I think, is relatively unproblematic (though by no means self-evident). The question of premise 3 is rather trickier.

We know by definition that intrinsic content properties must be inert in some important respects, and are quite limited in the ways they can relate to other items. The only examples of intrinsic contents that we are familiar with - the phenomenal properties of experiences - are certainly mental (or at least, if described systematically must turn out mental). These are promising signs, but fall a long way short of a complete proof.

Still, this is a project for another time and, perhaps, a larger body of work. For now we must content ourselves with the proof of reductive idealism about E, and one other intriguing fact. In advocating Foster's arguments, we showed that a key element in the constitution of the physical world is its *selection* as our world. The fact that, out of all the possible physical worlds causally deriving from F[156], E is *our* world turns out to be of paramount importance.

And there's a link here to special relativity - in which the spatial and temporal framework we use to interact with F is selected by facts about our state as observers - specifically, our velocity. Other observers located very close to us but moving at very different velocities will see a very different universe. Though there isn't time to go into the details here, selection also crops up as a key feature of quantum physics.

It would be trite and cheap to leap straight to the claim that only a conscious or mental phenomenon could stand behind all this selection. If nothing else, the scientific theories are - as most scientists would acknowledge - far from complete. The logic we must use to engage with them - which still fundamentally emerges out of our perceptual paradigm, and is hard-pressed to handle even basic phenomena like time dilation and order switching - gets murky when pushed to such extremes.

---

[156] and it's worth noting that other creatures may be inhabiting their own experiential worlds derived from the same F but utterly different from E, purely in virtue of the fact that they relate to F in a way significantly different to ours.

Still, we can at least give our question something of a scientific framework using the concept of selection. Let us assume - and many scientists are engaged in a frantic struggle to disprove this assumption, a struggle which may yet bear fruit - that selection is a necessary feature of all the physics that it currently seems to be a part of[157]. So, for each spatial and temporal interval, for each quantum state and so on, there must be some selector.

In many if not most cases, those selectors may turn out to be other physical phenomena, themselves selected by some ontologically (or perhaps just 'selection-wise', if you'll forgive the clumsy term) prior phenomenon. But if we assume that some states are selected - and they certainly appear to be, which may in this case be the same as saying that they are - from the manifold of possibility, then somewhere at the bottom of all this there must be a non-selected selector.

Perhaps this sounds a little too much like a first cause argument. Bear with me; I do not claim to be able to make such an argument. I claim only that my understanding of our best current scientific theories and ordinary experience matches up to these two premises:

1. The putative fundamental entities of physics all require selection.
2. Some physical states are selected.

These two put together require that there be some non-physical entity to perform the initial selection that drags the (our) physical world into actuality. But by our terms, everything that is non-physical is mental. And so everything physical will have a dependence on a mental selector.

This is, of course, a pitifully brief sketch of an argument. I do not claim it is correct. Instead, I bring it up to illustrate a particular key point: the realist can try to block the path to full-blooded idealism by finding some physical postulate which does not require selection and is fitted to serve as a selector. Conversely, in some ways the idealist's best hope is to shore up premise 1 above: to show that all physical fundamentals require selection.

In this way, the conflict between realism and idealism may yet turn out to have its final act on the scientific rather than the philosophical stage; it is even conceivable that the two may merge back together. Either way, it will be interesting to find out.

---

[157] Particularly in the field of quantum physics, there are more and less egregiously non-physical construals of the selection process. Those I am aware of, however, all feature some sort of selection - they make the features of the physical world (as we perceive it) depend on the relationship of that physical world to us.

Bibliography

Abbott, E.A. 1998. Flatland: A Romance of Many Dimensions. London: Penguin Books

Aloni, M. 2005. A Formal Treatment of the Pragmatics of Questions and Attitudes. *Linguistics and Philosophy*, [e-journal] 28(5) pp.505-539. Available through: Springer Link website <link.springer.com> [last accessed 27th August 2013]

Bars, I. 2006. The Standard Model of Particles and Forces in the Framework of 2T-physics. *Physical Review D*, [e-journal] 74(8). Available through: American Physical Society website <prd.aps.org> [last accessed 27th August 2013]

Benesty, J. Sondhi, M.M. and Huang, Y.A. eds. 2008. *Springer Handbook of Speech Processing*. [e-book] Heidelberg: Springer Science & Business Media. Available through: Liverpool University Library website <library.liv.ac.uk> [last accessed 27th August 2013]

Bird, A. 2007a. *Nature's Metaphysics*. Oxford: Oxford University Press

Bird, A. 2007b. The Regress of Pure Powers? *The Philosophical Quarterly*, [e-journal] 57(229) pp.513-534. Available through: Wiley Online Library website <onlinelibrary.wiley.com> [last accessed 21st September 2013]

Bradley, F.H., 1908. *Appearance and Reality*. 2nd ed. London: Swan Sonnenschein & co.

Buroker, J.V. 2006. *Kant's 'Critique of Pure Reason'*. [e-book] Cambridge: Cambridge University Press. Available through: Liverpool University Library website <www.library.liv.ac.uk> [last accessed 27th August 2013]

Candlish, S. 2007. *The Russell/Bradley Dispute and its Significance for Twentieth-Century Philosophy*. Basingstoke: Palgrave Macmillan

Cox, B. and Forshaw, J. 2009. *Why Does $E=mc^2$?* Cambridge, MA: Da Capo Press

Dipert, R. 1997. The Mathematical Structure of the World: The World as Graph. *The Journal of Philosophy*, [e-journal] 94(7) pp.329-358. Available through: JSTOR website <www.jstor.org> [last accessed 27th August 2013]

Einstein, A. 1916. *Relativity: The Special and General Theory*. [e-book] Translated by R.W. Lawson. Available through: Wikisource website <en.wikisource.org/wiki> [last accessed 27th August 2013]

Feigl, H. 1967. *The 'Mental' and the 'Physical': the essay and a postscript*. Minneapolis: University of Minnesota Press

Foster, J. 1982. *The Case for Idealism*. London: Routledge & Kegan Paul

Foster, J. 1991. *The Immaterial Self*. London: Routledge

Foster, J. 2008. *A World for Us*. Oxford: Oxford University Press

Hylton, P. 1990. *Russell, Idealism and the Emergence of Analytic Philosophy*. Oxford: Clarendon Press

Kant, I. 1787. *Critique of Pure Reason*. Translated by P. Guyer and A.W. Wood, 1998. New York: Cambridge University Press

Lewis, D. 2009. *Ramseyan Humility*. In: D. Braddon-Mitchell and R. Nola, eds. 2009 *Conceptual Analysis and Philosophical Naturalism*. [e-book] Cambridge, MA: MIT Press. Available through: Liverpool University Library website <www.library.liv.ac.uk> [last accessed 27th August 2013]

Locke, D.T. 2008. Ramsification and Knowledge-Which. In: SEP (Society for Exact Philosophy), *36th Annual Meeting*. Laramie, WY May 13th-17th 2008.

Locke, D.T. 2009. A Partial Defence of Ramseyan Humility. In: D. Braddon-Mitchell and R. Nola, eds. 2009 *Conceptual Analysis and Philosophical Naturalism*. [e-book] Cambridge, MA: MIT Press. Available through: Liverpool University Library website <www.library.liv.ac.uk> [last accessed 27th August 2013]

National Research Council 2003. *The Polygraph and Lie Detection*. [e-book] Washington DC: The National Academies Press. Available through: National Academies Press website <nap.edu/catalog.php> [last accessed 27th August 2013]

Nicholson, P.P. ed. 1997. *Collected Works of T.H. Green*. Bristol: Thoemmes

Oderberg, D.S. 2011. The World is not an Asymmetric Graph. *Analysis*, [e-journal] 71(1) pp.3-10. Available through: Oxford Journals website <analysis.oxfordjournals.org> [last accessed 27th August 2013]

Oxford 2013. *Oxford Dictionaries Online*. [online](2013) Available at: <oxforddictionaries.com> [last accessed 27th August 2013]

Quinton, A. 1962. Spaces and Times. *Philosophy*, [e-journal] 37(140) pp.130-147. Available through: Cambridge Journals Online <journals.cambridge.org> [last accessed 27th August 2013]

Van Rooy, R. 2003. Questioning to Resolve Decision Problems. *Linguistics and Philosophy*, [e-journal] 26(6) pp.727-763. Available through: Springer Link website <link.springer.com> [last accessed 27th August 2013]

Schaffer, J. 2005. Quiddistic Knowledge. *Philosophical Studies*, [e-journal] 123(1-2), pp.1-32. Available through: Springer Link website <link.springer.com> [last accessed 27th August 2013]

Shabel, L. 2010. The Transcendental Aesthetic. In: P. Guyer, ed. 2010. *The Cambridge Companion to Kant's Critique of Pure Reason*. [e-book] New York: Cambridge University Press. Available through: Liverpool University Library website <library.liv.ac.uk> [last accessed 27th August 2013]

't Hooft, G. 1993. *Dimensional Reduction in Quantum Gravity*. [online] Available through: arxiv.org website <http://arxiv.org/abs/gr-qc/9310026> [last accessed 21st September 2013]

Tomassi, P. 1999. *Logic*. London; New York: Routledge